

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:55:59 ; Search time 2988 Seconds  
(without alignments)  
18293.095 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagttgcgcgtgcac.....tgtcaccttgaggtcttc 3375

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estnu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_htc:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_htc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	305	9.0	535	12	BF849816
c 2	296	8.8	406	10	AW206492
c 3	128	3.8	282	10	BE143292
c 4	113	3.3	723	14	BM702190
5	69	2.0	496	17	AQ202422
6	59	1.7	567	10	AW855319

c	7	1.7	201	10	AW856235
	31	0.9	631	10	BB612001
	9	0.9	638	10	BB613951
	10	0.9	661	10	BB617396
	11	0.9	664	10	BB617396
	12	0.9	661	12	BF601265
	13	0.9	148	13	BG991167
	29	0.7	395	9	AA910328
c	13	0.7	445	9	AI744019
c	14	0.7	474	9	AI793020
c	15	0.7	494	9	AI373743
c	16	0.7	557	10	AW241667
c	17	0.7	348	9	AI772957
c	18	0.7	519	9	AI018418
c	19	0.7	586	10	AW648529
c	20	0.7	204	10	BB571653
c	21	0.6	285	17	BH273274
c	22	0.6	300	9	AI991076
c	23	0.6	371	10	BE366992
c	24	0.6	371	17	AQ004063
c	25	0.6	469	9	AA103567
c	26	0.6	480	10	BE367090
c	27	0.6	490	9	AI668711
c	28	0.6	495	17	B50693
c	29	0.6	622	9	AI054920
c	30	0.6	632	17	A2344444
c	31	0.6	683	17	AQ957276
c	32	0.6	748	14	BM963047
c	33	0.6	750	14	BM947057
c	34	0.6	789	14	BQ443279
c	35	0.6	930	17	AQ749010
c	36	0.6	936	14	BQ941497
c	37	0.6	120	13	BM484625
c	38	0.6	194	9	AA078411
c	39	0.6	197	12	BF461151
c	40	0.6	197	12	BF461673
c	41	0.6	241	9	AA767572
c	42	0.6	241	17	AZ744203
c	43	0.6	256	10	AV361506
c	44	0.6	277	10	AW516927
c	45	0.6			

#### ALIGNMENTS

#### RESULT 1

BF849816/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BF849816 535 bp mRNA linear EST 16-JAN-2001  
PM4-EN0068-151100-004-b06 EN0068 Homo sapiens cDNA, mRNA sequence.

BF849816

BF849816.1 GI:12236966

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 535)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001



Qy 2856 CCACAACGATGTTGTGAAAGTTTGTGATGTGTAGTAATAACCCACTTCTCTA 2911  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 72 CCACAACGATGTTGTGAAAGTTTGTGATGTGTAGTAATAACCCACTTCTCTA 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 3  
 BE143292/c  
 LOCUS BE143292 282 bp mRNA linear EST 21-JUN-2000  
 DEFINITION MR0-HT0161-221099-002-c08 HT0161 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE143292  
 VERSION BE143292.1 GI:8606013  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 282)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.J.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-MR0-HT0161-221  
 099-002-c08&t3=1999-10-22&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 8  
 High quality sequence stop: 282.

FEATURES  
 source  
 1..282  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0161"  
 /dev\_stage="Adult"  
 /note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 71 a 78 c 63 g 70 t  
 ORIGIN  
 Query Match 3.8%; Score 128; DB 10; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-52;  
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3026 CGCCAGATGTTGTGACACAGTAATGCCAGCAGAGCGCTTTACTAGAGCATCCTTTGG 3085  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 282 CGCCAGATGTTGTGACACAGTAATGCCAGCAGAGCGCTTTACTAGAGCATCCTTTGG 223  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 3086 ACGCGGAAGCCAGCGCTTTCAGATGGAAGCAGCAGCTTTTCCACTTCCCGCAGAC 3145  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 222 ACGCGGAAGCCAGCGCTTTCAGATGGAAGCAGCAGCTTTTCCACTTCCCGCAGAC 163  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 3146 ATTCTGGA 3153  
 ||||||||  
 Db 162 ATTCTGGA 155  
 ||||||||

RESULT 4  
 BM702190/c  
 LOCUS BM702190 723 bp mRNA linear EST 28-FEB-2002  
 DEFINITION UI-E-CQ1-aeY-m-08-0-UI.r1 UI-E-CQ1 Homo sapiens cDNA clone  
 UI-E-CQ1-aeY-m-08-0-UI 5', mRNA sequence.  
 ACCESSION BM702190  
 VERSION BM702190.1 GI:19015448  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 723)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 353-608, >LINE2  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..723  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-CQ1-aeY-m-08-0-UI"  
 /clone\_lib="UI-E-CQ1"  
 /tissue\_type="optic nerve"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: pT73-pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 UI-E-CQ1 is a normalized cDNA library containing the  
 following tissue(s): optic nerve. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 CCATTAAGTC. This library was created for the program, Gene  
 Discovery in the Visual System, supported by National Eye  
 Institute (NEI)."  
 BASE COUNT 147 a 241 c 142 g 190 t 3 others  
 ORIGIN  
 Query Match 3.3%; Score 113; DB 14; Length 723;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-45;  
 Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2615 GGATTTCTGAGAGCCCTCGAGGCATGCTCCCTGTCAGGAGGGCAGCAGCGGTACCC 2674  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 113 GGATTTCTGAGAGCCCTCGAGGCATGCTCCCTGTCAGGAGGGCAGCAGCGGTACCC 54  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 2675 CTCCCAGCACTACAGAGAGCGCTGGGCGACTGAAATGGTGCTACCTTCTGG 2727  
 ||||||||







```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="5031412M06"
/clone_lib="RIKEN full-length enriched, 11 days pregnant
adult female ovary and uterus"
/sex="female"
/tissue_type="ovary and uterus"
/dev_stage="11 days pregnant, adult"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGAGCTCTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAAATTAATTCCTCCCTCCCTCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI."
BASE COUNT      134 a 196 c 195 g 139 t
ORIGIN
Query Match      0.9%; Score 31; DB 10; Length 664;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 GTCCACTCTCATCTGGGATTCCTCGA 644
|||||
Db 363 GTCCACTCTCATCTGGGATTCCTCGA 393

RESULT 11
BF601265      501 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Cassas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keefe, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel.: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT

```

```

BACKWARD: GTTTCCTCCAGTCAGCAGC
Plate: 40 row: J column: 10
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..501
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: NotI; Site_2: Sali;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT      98 a 164 c 154 g 85 t
ORIGIN
Query Match      0.9%; Score 30; DB 12; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2547 TGGCTGGAGGGCCCCCTGCGAGAACCG 2576
|||||
Db 24 TGGCTGGAGGGCCCCCTGCGAGAACCG 53

RESULT 12
BG991167      148 bp mRNA linear EST 13-JUN-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel.: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR26t2-MR2-
270101-003-h07&t3=2001-01-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 33
High quality sequence stop: 147.
Location/Qualifiers
1..148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT1189"
/dev_stage="Adult"
/notes="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)

```

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 29 a 43 c 47 g 29 t

ORIGIN

Query Match 0.7%; Score 29; DB 13; Length 148;  
Best Local Similarity 100.0%; Pred. No. 0.0064;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2743 GTCCTTAGAATGCTGCTCCGCGGTG 2771  
|||||  
Db 42 GTCCTTAGAATGCTGCTCCGCGGTG 70  
|||||

RESULT 13  
AA910328/c  
LOCUS AA910328 395 bp mRNA linear EST 13-APR-1998  
DEFINITION ok83c11.s1 NCI\_CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:1520564 3'  
similar to TR:Q23242 Q23242 ZC116.3 ; mRNA sequence.  
ACCESSION AA910328  
VERSION AA910328.1 GI:3049618  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 395)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: 40m13 fwd EF from Amersham.  
Location/Qualifiers  
1. 395

FEATURES source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1520564"  
/clone\_lib="NCI\_CGAP\_Kid3"  
/lab\_host="DH10B"  
/note="organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 90 a 101 c 84 g 120 t

ORIGIN

Query Match 0.7%; Score 25; DB 9; Length 395;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1236 CTGCCAGATGGAGGCACATGTGTT 1260  
|||||  
Db 333 CTGCCAGATGGAGGCACATGTGTT 309  
|||||

RESULT 14  
AA1744019/c

LOCUS AA1744019 445 bp mRNA linear EST 17-DEC-1999  
DEFINITION wc32g02.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2316914 3'  
similar to TR:O60494 O60494 INTRINSIC FACTOR-B12 RECEPTOR  
PRECURSOR. ; mRNA sequence.  
ACCESSION AA1744019  
VERSION AA1744019.1 GI:5112307  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 445)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 589 Std Error: 0.00  
Seq primer: 40UP from Gibco.  
Location/Qualifiers  
1. 445

FEATURES source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2316914"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5 000 clones made from the same library (cloneids 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 107 a 111 c 95 g 132 t

ORIGIN

Query Match 0.7%; Score 25; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 0.77;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1236 CTGCCAGATGGAGGCACATGTGTT 1260  
|||||  
Db 344 CTGCCAGATGGAGGCACATGTGTT 320  
|||||

RESULT 15  
AA1793020

LOCUS AA1793020 474 bp mRNA linear EST 13-DEC-1999  
DEFINITION qz39a03.y5 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2029228 5'  
similar to TR:O60494 O60494 INTRINSIC FACTOR-B12 RECEPTOR  
PRECURSOR. ; mRNA sequence.  
ACCESSION AA1793020  
VERSION AA1793020.1 GI:5340736  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 474)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

FEATURES source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2316914"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5 000 clones made from the same library (cloneids 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 107 a 111 c 95 g 132 t

ORIGIN

Query Match 0.7%; Score 25; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 0.77;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1236 CTGCCAGATGGAGGCACATGTGTT 1260  
|||||  
Db 344 CTGCCAGATGGAGGCACATGTGTT 320  
|||||

RESULT 15  
AA1793020

LOCUS AA1793020 474 bp mRNA linear EST 13-DEC-1999  
DEFINITION qz39a03.y5 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2029228 5'  
similar to TR:O60494 O60494 INTRINSIC FACTOR-B12 RECEPTOR  
PRECURSOR. ; mRNA sequence.  
ACCESSION AA1793020  
VERSION AA1793020.1 GI:5340736  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 474)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

FEATURES source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2316914"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5 000 clones made from the same library (cloneids 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 107 a 111 c 95 g 132 t

ORIGIN

Query Match 0.7%; Score 25; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 0.77;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1236 CTGCCAGATGGAGGCACATGTGTT 1260  
|||||  
Db 344 CTGCCAGATGGAGGCACATGTGTT 320  
|||||

RESULT 15  
AA1793020

LOCUS AA1793020 474 bp mRNA linear EST 13-DEC-1999  
DEFINITION qz39a03.y5 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2029228 5'  
similar to TR:O60494 O60494 INTRINSIC FACTOR-B12 RECEPTOR  
PRECURSOR. ; mRNA sequence.  
ACCESSION AA1793020  
VERSION AA1793020.1 GI:5340736  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 474)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
COMMENT  
Tumor Gene Index  
Unpublished (1997)  
Other ESTs: qz39a03.xl  
Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-re@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cdNA Library Preparation: M. Bento Soares, Ph.D.  
cdNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.ntml

This read is a RESSEQUENCE of a previously sequenced human clone  
Original clone citation: see original entry for original citation  
information  
This 5' resequenced clone has no previous 5' data to verify this  
new read against  
Insert Length: 878 Std Error: 0.00  
Seq primer: -40RP from Gibco  
High quality sequence stop: 425.

FEATURES  
Location/Qualifiers  
1..474  
Source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2029228"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 148 a 89 c 107 g 130 t  
ORIGIN  
Query Match 0.78; Score 25; DB 9; Length 474;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1236 CTGCCAGATGGAGCACATGTGT 1260  
Db 411 CTGCCAGATGGAGCACATGTGT 435  
|||||

Search completed: May 5, 2003, 07:32:42  
Job time : 3012 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 22:50:07 ; Search time 5749 Seconds  
(without alignments)  
17085.053 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgttcgcgcgcac.....tgtccacctgaaggtcttc 3375

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	712.6	21.1	123110	9	AC005383	AC005383 Homo sapi
2	408.6	12.1	152895	2	AC119647	AC119647 Rattus no
3	408.6	12.1	154748	2	AC125150	AC125150 Mus muscu
C 4	317.8	9.4	160420	9	AC022023	AC022023 Homo sapi
C 5	309.8	9.2	136357	2	AL161942	AL161942 Homo sapi
C 6	123.4	3.7	65824	2	AC099899	AC099899 Mus muscu
C 7	111.8	3.3	65824	2	AC099899	AC099899 Mus muscu
C 8	97.8	2.9	4151	6	AX281619	AX281619 Sequence
C 9	96	2.8	152895	2	AC119647	AC119647 Rattus no
C 10	95.6	2.8	1304	5	CHKMP	M14792 Chicken car
C 11	95.6	2.8	3327	9	HUMCMPMR	N55683 Human carti
C 12	87.8	2.6	7504	9	AB040943	AB040943 Homo sapi
C 13	86.8	2.6	194143	2	AC112730	AC112730 Rattus no
C 14	85	2.5	1959	10	MMU35035	U35035 Mus musculu
C 15	82	2.4	190669	10	AL590429	AL590429 Mouse DNA
C 16	80	2.4	3571	10	MMU69262	U89262 Mus musculu
C 17	79.4	2.4	9235	10	MMU32107	U32107 Mus musculu
C 18	78.6	2.3	2193	10	MMU6140	AJ006140 Mus muscu
C 19	78.4	2.3	3273	10	BC005429	BC005429 Mus muscu
C 20	77	2.3	169585	2	AC078821	AC078821 Homo sapi
C 21	76.6	2.3	186666	2	AC097136	AC097136 Rattus no
C 22	76.4	2.3	177695	2	AC096154	AC096154 Rattus no
C 23	76.4	2.3	192817	2	AC127063	AC127063 Rattus no
C 24	75.4	2.2	1360	9	HUMCMP5	M55679 Human carti
C 25	75.4	2.2	152042	9	AL137857	AL137857 Human DNA
C 26	75.2	2.2	3550	9	BC010444	BC010444 Homo sapi
C 27	74.8	2.2	202006	2	AL669980	AL669980 Mus muscu
C 28	74.2	2.2	1167	9	AK027323	AK027323 Homo sapi
C 29	74.2	2.2	2003	9	HS47581	AJ007581 Homo sapi
C 30	73.6	2.2	1033	6	AX329755	AX329755 Sequence
C 31	73.6	2.2	2293	9	AK027775	AK027775 Homo sapi
C 32	73.6	2.2	2733	9	BC016394	BC016394 Homo sapi
C 33	73.6	2.2	2746	9	HSM802406	AL137638 Homo sapi
C 34	73.6	2.2	2863	6	AX015418	AX015418 Sequence
C 35	73.6	2.2	3032	6	AX399978	AX399978 Sequence
C 36	73.6	2.2	3089	6	AX399977	AX399977 Sequence
C 37	73.6	2.2	3373	6	AR085071	AR085071 Sequence
C 38	73.6	2.2	3476	6	AX079878	AX079878 Sequence
C 39	73.6	2.2	3449	6	AX464160	AX464160 Sequence
C 40	73.6	2.2	3496	9	HSU69263	U69263 Homo sapien
C 41	73.6	2.2	3572	9	HSM805245	AL833931 Homo sapi
C 42	73.6	2.2	147620	9	HS453C12	AL021578 Human DNA
C 43	73.6	2.2	189656	9	AF314058	AF314058 Homo sapi
C 44	73.4	2.2	595	5	GCGMP5	X12350 Chicken car
C 45	72.6	2.2	3026	6	AX079876	AX079876 Sequence

ALIGNMENTS

RESULT 1  
AC005383  
LOCUS AC005383 123110 bp DNA linear PRI 31-OCT-1998  
DEFINITION Homo sapiens chromosome 10 clone C1987SK-114466 map 10q25.1,  
complete sequence.  
ACCESSION AC005383  
VERSION AC005383.1 GI:3818355  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 123110)  
AUTHORS Smith,D.R.  
TITLE Sequencing of Human Chromosome 10

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 123110)  
AUTHORS Smith, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (05-AUG-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REFERENCE 3 (bases 1 to 123110)  
AUTHORS Smith, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (31-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

COMMENT On Oct 31, 1998 this sequence version replaced gi:3808081.

FEATURES  
source  
Location/Qualifiers  
1..123110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.1"  
/clone="C1987SK-1144G6"  
BASE COUNT 31394 a 28668 c 29634 g 33413 t 1 others  
ORIGIN

Query Match 21.1%; Score 712.6; DB 9; Length 123110;  
Best Local Similarity 98.4%; Pred. No. 4.2e-141;  
Matches 751; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

QY 2615 GGATCTTGAGAGCGCCCTGAGGCATGCTCCCGTGCAGGAGGCGAGCGGTACCC 2674  
Db 78153 GGATCTTGAGAGCGCCCTGAGGCATGCTCCCGTGCAGGAGGCGAGCGGTACCC 78212

QY 2675 CTCCAGCACTACAGAGAGCGCTGGCACTGAATGTGGCTTCTGCAATGTCT 2734  
Db 78213 CTCCAGCACTACAGAGAGCGCTGGCACTGAATGTGGCTTCTGCAATGTCT 78272

QY 2735 GTCCCGCAGCTCTAGAAATGTCTGCTTCCCGCGCTGGCCAGGACCACTATTCTCACTGA 2794  
Db 78273 GTCCCGCAGCTCTAGAAATGTCTGCTTCCCGCGCTGGCCAGGACCACTATTCTCACTGA 78332

QY 2795 GGAGGAGGATGCCCACTGCAGCCATGCTGCTTAGAGACAAGAGAGCTGATGTCA 2854  
Db 78333 GGAGGAGGATGCCCACTGCAGCCATGCTGCTTAGAGACAAGAGAGCTGATGTCA 78392

QY 2855 CCACAAACATGTGTGTTGAAAAGTTTGTATGTAGTAAATACCCACTTTCTGTACCT 2914  
Db 78393 CCACAAACATGTGTGTTGAAAAGTTTGTATGTAGTAAATACCCACTTTCTGTACCT 78452

QY 2915 GCTGTGCTTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAGGGTC 2974  
Db 78453 GCTGTGCTTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAGGGTC 78512

QY 2975 CTGAAGACTTAAATTTAGCGGCTGAGTTCCTTTGACACAATCAATGCTGCCAGAT 3034  
Db 78513 CTGAAGACTTAAATTTAGCGGCTGAGTTCCTTTGACACAATCAATGCTGCCAGAT 78572

QY 3035 GTTGTTCACACAGTAATGCCACAGAGCGCTTTACTAGAGCATCTTTTGAGCGGCGAAG 3094  
Db 78573 GTTGTTCACACAGTAATGCCACAGAGCGCTTTACTAGAGCATCTTTTGAGCGGCGAAG 78632

QY 3095 GCCACGCGCTTTCAAGATGGAAGACAGCAGCTTTTCCACATCTCCCGAGACATCTTGGAT 3154  
Db 78633 GCCACGCGCTTTCAAGATGGAAGACAGCAGCTTTTCCACATCTCCCGAGACATCTTGGAT 78692

QY 3155 GCATTTGCATGAGTCTGAAGGGGGTCTGAGGAGCTTTGTGACTTCTTTGGGCACTGCC 3214  
Db 78693 GCATTTGCATGAGTCTGAAGGGGGTCTGAGGAGCTTTGTGACTTCTTTGGGCACTGCC 78751

QY 3215 TTTTGTGCTGGAGAGACTTGGAAAGGTCTCAGACTG-AATGTGACCAATTAAACAGCT 3273  
Db 78752 TTTTGTGCTGGAGAGACTTGGAAAGGTCTCAGACTGAAATGTGACCAATTAAACAGCT 78811

QY 3274 TGGTTGATGATGGGGGAGGGGCTGAGTTGTGCATGGCCCGCCAGGTCTGGA-GGGCCACGTA 3332  
Db 78812 TGGTTGATGATGGGGGAGGGGCTGAGTTGTGCATGGCCCGCCAGGTCTGGAAGGGCCACGTA 78871

QY 3333 AAATCTTTCTGAGTCTGAGCAGTGTCCACCTTGAAGGTCTTC 3375  
Db 78872 AAATCTTTCTGAGTCTGAGCAGTGTCCACCTTGAAGGTCTTC 78914

RESULT 2  
AC119647

LOCUS AC119647 152895 bp DNA linear HTG 18-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-137H14, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 54 unordered pieces.

ACCESSION AC119647  
VERSION AC119647.2 GI:21747172  
KEYWORDS HTG; HTGS-PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 152895)

AUTHORS

Muzdy, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbarta, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, F.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davella, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Dunaway, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Foster, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Homi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, F., Johnson, R., Jolivet, S., Joudan, S.,  
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, B., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwono, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sissosh, I.,  
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

TITLE Direct Submission

JOURNAL

REFERENCE 2 (bases 1 to 152895)

AUTHORS

Worley, K.C.

TITLE Direct Submission

JOURNAL

Submitted (30-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 152895)

AUTHORS

Worley, K.C.

TITLE Direct Submission

JOURNAL

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One





Db 108525 GGCACCACTGGGGGCGCTCCGGAGGCCCAAGCCCTTTGCTCAAGCCCTTTGTGTGCAGGCC 108584

Qy 1444 GTGCTGAGCAGGAGACTCTCGGGGCCGAGTGGGTGTGCCACATACAGCAGGAGGCTGCTG 1503

Db 108585 GTGCTGAGGAGGAGACTCCCGAGCCCGCTTGGGATAGCCAGTTATGCGAGGAATCTAATG 108644

Qy 1504 GTGCGGTGCTGTGGGGGAGTACCAAGGATGCGCTGACCTGCTGTGAGCCCTCGATGGC 1563

Db 108645 GTGCGGTGCCCCGTGGGGGAGTACCAACATGTGCGGACCTGATCAGGAGCCTTGACAGC 108704

Qy 1564 ATTCCCTTCGCTGCTGTGCCGCCACCCCTGACGGGAGTGCCTTGGCGGAGCGGCAGAGCGT 1623

Db 108705 ATTCCCTTCAGGCTGSCCCGACCCCTAACCGGAGTGCCTTGTCTCCAGGTGGCAGAGCAC 108764

Qy 1624 GGCTTCGGGAGGCCACACAGGACAGGCCAGGCCGCGCCACGTAGAGTGGTGGTTTGCTC 1683

Db 108765 GGCTTTGGGAGTGCCACGACGAGCTGGTCAGGACAGGCCACGCGAGATGATGTTCTGTCTC 108824

Qy 1684 ACTGAGTGCACACTCCGAGGATGAGTTTGGGGGCCACGCGCTCACGCAAGGGCGCCAGAG 1743

Db 108825 ACTGAATCACGCTCCCAAGGATGAGTGTCTTGGGCCACGAGCTCACGCAAGSGCTCGGGAG 108884

Qy 1744 CTGCTCTCTGTGGGTGTAGCAGATGAGCCGTGGGGCCAGCTGGAGGAGATCACAGGC 1803

Db 108885 CTACTCTCTGTGGCGTGGGCACTGAGATCCTGCAGCGGAGCTGGTGTGAAGATCACCGGT 108944

Qy 1804 AGCCCAAGCATGTGATGTGCTACTCGGATCCTCAGGATCTGTTCAACCAAAATCCCTGAG 1863

Db 108945 AGCCCCAAGCATGTGATGGTTCACACAGACCCCTCAGGACCTGTTCCAGCCAAATCCCNAG 109004

Qy 1864 CTGCGGGGAGCTGTGCAGCCGGCAGCGGCCAGG 1898

Db 109005 CTGCAGAGGAGGCTATGCAGCCAGCCAGCCAGCCAGG 109039

```

RESULT 4
AC022023/c
LOCUS      AC022023      160420 bp      DNA      linear      PRI 17-MAY-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-411P18, complete sequence.
ACCESSION  AC022023
VERSION    AC022023.7  GI:20901843
KEYWORDS  HTG.
SOURCE     Homo sapiens.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith,D.R.
^Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 160420)
Smith,D.R.
Direct Submission
Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
3 (bases 1 to 160420)
Smith,D.R.
Direct Submission
Submitted (22-JUN-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
4 (bases 1 to 160420)
Smith,D.R.
Direct Submission
Submitted (17-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 17, 2002 this sequence version replaced gi:14522958.
Location/Qualifiers
1..160420
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-411P18"
FEATURES             source

```

BASE COUNT	48686 a	32379 c	32258 g	47097 t	
ORIGIN					
Query Match	9.4%	Score 317.8;	DB 9;	Length 160420;	
Best Local Similarity	95.1%	Pred. No. 3.8e-57;			
Matches 328;	Conservative 0;	Mismatches 17;	Indels 0;	Gaps 0;	
QY	2	ACAGTGTTCCGGCGTGCACCGCTCGGAGGCTGGGTGCACCGCTAGAGTCAAGTACTTTT	61		
Db	5373	ACAGTGTTCCGGCGTGCACCGCTCGGAGGCTGGGTGCACCGCTAGAGTCAAGTACTTTT	5314		
QY	62	TTTATTTTCAGACCTCGGGCCGATCCCGCTTTTAAAAACGCGAGGGGCTCTATGCACCTCC	121		
Db	5313	TTTATTTTCAGACCTCGGGCCGATCCCGCTTTTAAAAACGCGAGGGGCTCTATGCACCTCC	5254		
QY	122	CTGGCGGTAGTTCCTCGACCTCAGCCGGTGCGGGTGTCGGCCGCTCTCCAGCAGAGA	181		
Db	5253	CTGGCGGTAGTTCCTCGACCTCAGCCGGTGCGGGTGTCGGCCGCTCTCCAGCAGAGA	5194		
QY	182	CAACAGGTGTCACAGTGGCAGCGCCGCGCGCCCTCTGTGATCCCGTAGCGC	241		
Db	5193	CAACAGGTGTCACAGTGGCAGCGCCGCGCGCCCTCTGTGATCCCGTAGCGC	5134		
QY	242	CCCTCGGCCCGAGCGCGCGCGGCTCTGTAGTAGAGCGCCCGGACCCGAGCGCTGGT	301		
Db	5133	CCCTCGGCCCGAGCGCGCGCGGCTCTGTAGTAGAGCGCCCGGACCCGAGCGCTGGT	5074		
QY	302	CGCCGCTCTCTCCGTTATATCAACATGCCCCCTTTCCCTGTTCG	346		
Db	5073	CGCCGCTCTCTCCGTTATATCAACATGCCCCCTTTCCCTGTTCG	5029		
RESULT 5					
AL161942/c					
LOCUS	AL161942	136357 bp	DNA	linear	HTG 13-JUN-2001
DEFINITION	Homo sapiens chromosome 20 clone RP11-83B5, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.				
ACCESSION	AL161942				
VERSION	AL161942.5	GI:9931708			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_CANCELLED.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 136357)				
TITLE	Burton, J.				
JOURNAL	Direct Submission				
COMMENT	Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Aug 28, 2000 this sequence replaced gi:9863622.				

```

RESULT 5
AL161942/c
LOCUS      AL161942                136357 bp    DNA        linear    HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 20 clone RP11-83B5, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION  AL161942
VERSION    AL161942.5    GI:9931708
KEYWORDS   HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 136357)
AUTHORS    Burton, J.
TITLE      Direct Submission
JOURNAL    Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Aug 28, 2000 this sequence version replaced gi:9863622.
            ----- Genome Center
            Center: Sanger Centre
            Center 'code': SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: BA83B5
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: dye-terminator Big Dye; 100% of reads
            Consensus quality: 114775 bases at least Q40
            Consensus quality: 123431 bases at least Q30
            Consensus quality: 128697 bases at least Q20
            Insert size: 133157; sum-of-contigs
            Quality coverage: 2.20x in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as

```

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 5425: contig of 5425 bp in length  
5426 5525: gap of 100 bp  
5526 9483: contig of 3958 bp in length  
9484 9583: gap of 100 bp  
9584 14280: contig of 4697 bp in length  
14281 14380: gap of 100 bp  
14381 17312: contig of 2932 bp in length  
17313 17412: gap of 100 bp  
17413 22848: contig of 5436 bp in length  
22849 22948: gap of 100 bp  
22949 26090: contig of 3142 bp in length  
26091 26190: gap of 100 bp  
26191 28215: contig of 2025 bp in length  
28216 28315: gap of 100 bp  
28316 33862: contig of 5547 bp in length  
33863 33962: gap of 100 bp  
33963 44145: contig of 10183 bp in length  
44146 44245: gap of 100 bp  
44246 53165: contig of 8920 bp in length  
53166 53265: gap of 100 bp  
53266 57412: contig of 4147 bp in length  
57413 57512: gap of 100 bp  
57513 60170: contig of 2658 bp in length  
60171 60270: gap of 100 bp  
60271 65985: contig of 5715 bp in length  
65986 66085: gap of 100 bp  
66086 69094: contig of 3009 bp in length  
69095 69194: gap of 100 bp  
69195 71736: contig of 2542 bp in length  
71737 71836: gap of 100 bp  
71837 74955: contig of 3119 bp in length  
74956 75055: gap of 100 bp  
75056 77325: contig of 2270 bp in length  
77326 77425: gap of 100 bp  
77426 80709: contig of 3284 bp in length  
80710 80809: gap of 100 bp  
80810 82943: contig of 2134 bp in length  
82944 83043: gap of 100 bp  
83044 85218: contig of 2175 bp in length  
85219 85318: gap of 100 bp  
85319 89861: contig of 4543 bp in length  
89862 89961: gap of 100 bp  
89962 92824: contig of 2863 bp in length  
92825 92924: gap of 100 bp  
92925 95822: contig of 2898 bp in length  
95823 95922: gap of 100 bp  
95923 98130: contig of 2208 bp in length  
98131 98230: gap of 100 bp  
98231 100594: contig of 2364 bp in length  
100595 100694: gap of 100 bp  
100695 109043: contig of 8349 bp in length  
109044 109143: gap of 100 bp  
109144 112146: contig of 3003 bp in length  
112147 112246: gap of 100 bp  
112247 117183: contig of 4937 bp in length  
117184 117283: gap of 100 bp  
117284 121861: contig of 4578 bp in length  
121862 121961: gap of 100 bp  
121962 124258: contig of 2297 bp in length  
124259 124358: gap of 100 bp  
124359 127036: contig of 2678 bp in length  
127037 127136: gap of 100 bp  
127137 133357: contig of 6221 bp in length  
133358 133457: gap of 100 bp  
133458 136357: contig of 2900 bp in length.

Location/Qualifiers  
1. .136357  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

FEATURES  
source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

/chromosome="20"  
/clone="RP11-83B5"  
/clone\_lib="RPC1-11.1"  
1. 5425  
/note="assembly\_fragment:01370"  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left  
5526..9483  
/note="assembly\_fragment:01120"  
fragment\_chain:1  
9584..14280  
/note="assembly\_fragment:00095"  
fragment\_chain:1  
14381..17312  
/note="assembly\_fragment:01155"  
fragment\_chain:1  
17413..22848  
/note="assembly\_fragment:00025"  
fragment\_chain:2  
22949..26090  
/note="assembly\_fragment:00387"  
fragment\_chain:2  
26191..28215  
/note="assembly\_fragment:00262"  
fragment\_chain:3  
28316..33862  
/note="assembly\_fragment:00134"  
fragment\_chain:3  
33963..44145  
/note="assembly\_fragment:00605"  
fragment\_chain:4  
44246..53165  
/note="assembly\_fragment:00989"  
fragment\_chain:4  
53266..57412  
/note="assembly\_fragment:00876"  
fragment\_chain:5  
57513..60170  
/note="assembly\_fragment:00248"  
fragment\_chain:5  
60271..65985  
/note="assembly\_fragment:01554"  
fragment\_chain:6  
66086..69094  
/note="assembly\_fragment:00470"  
fragment\_chain:6  
69195..71736  
/note="assembly\_fragment:01566"  
fragment\_chain:7  
71837..74955  
/note="assembly\_fragment:00477"  
fragment\_chain:7  
75056..77325  
/note="assembly\_fragment:00036"  
77426..80709  
/note="assembly\_fragment:00390"  
80810..82943  
/note="assembly\_fragment:00409"  
83044..85218  
/note="assembly\_fragment:00437"  
85319..89861  
/note="assembly\_fragment:00764"  
89962..92824  
/note="assembly\_fragment:00782"  
92925..95822  
/note="assembly\_fragment:00808"  
95923..98130  
/note="assembly\_fragment:00831"  
98231..100594  
/note="assembly\_fragment:00886"  
100695..109043  
/note="assembly\_fragment:00992"

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information -----

Center project name: L13492

Center clone name: 12\_A\_20

-----

\* NOTE: This record contains 82 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is complete.y

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1 689: contig of 689 bp in length

\* 690 789: gap of 100 bp

\* 790 1505: contig of 716 bp in length

\* 1506 1605: gap of 100 bp

\* 1606 2298: contig of 693 bp in length

\* 2299 2398: gap of 100 bp

\* 2399 3092: contig of 694 bp in length

\* 3093 3192: gap of 100 bp

\* 3193 3864: contig of 672 bp in length

\* 3865 3964: gap of 100 bp

\* 3965 4677: contig of 713 bp in length

\* 4678 4777: gap of 100 bp

\* 4778 5469: contig of 692 bp in length

\* 5470 5569: gap of 100 bp

\* 5570 6268: contig of 699 bp in length

\* 6269 6368: gap of 100 bp

\* 6369 7088: contig of 720 bp in length

\* 7089 7188: gap of 100 bp

\* 7189 7901: contig of 713 bp in length

\* 7902 8001: gap of 100 bp

\* 8002 8705: contig of 704 bp in length

\* 8706 8805: gap of 100 bp

\* 8806 9503: contig of 698 bp in length

\* 9504 9603: gap of 100 bp

\* 9604 10289: contig of 686 bp in length

\* 10290 10389: gap of 100 bp

\* 10390 11101: contig of 712 bp in length

\* 11102 11201: gap of 100 bp

\* 11202 11908: contig of 707 bp in length

\* 11909 12008: gap of 100 bp

\* 12009 12714: contig of 706 bp in length

\* 12715 12814: gap of 100 bp

\* 12815 13516: contig of 702 bp in length

\* 13517 13616: gap of 100 bp

\* 13617 14321: contig of 705 bp in length

\* 14322 14421: gap of 100 bp



Ginde,S., Gord,S., Goyette,M., Graham,L., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
 Meneus,L., Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Vial,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL  
 COMMENT

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L13492  
 Center clone name: 12\_A\_20

NOTE: This record contains 82 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 689: contig of 689 bp in length  
 \* 690 789: gap of 100 bp  
 \* 790 1505: contig of 716 bp in length  
 \* 1506 1605: gap of 100 bp  
 \* 1606 2298: contig of 693 bp in length  
 \* 2299 2398: gap of 100 bp  
 \* 2399 3092: contig of 694 bp in length  
 \* 3093 3192: gap of 100 bp  
 \* 3193 3864: contig of 672 bp in length  
 \* 3865 3964: gap of 100 bp  
 \* 3965 4677: contig of 713 bp in length  
 \* 4678 4777: gap of 100 bp  
 \* 4778 5469: contig of 692 bp in length  
 \* 5470 5569: gap of 100 bp  
 \* 5570 6268: contig of 699 bp in length  
 \* 6269 6368: gap of 100 bp  
 \* 6369 7088: contig of 720 bp in length  
 \* 7089 7188: gap of 100 bp  
 \* 7189 7901: contig of 713 bp in length  
 \* 7902 8001: gap of 100 bp  
 \* 8002 8705: contig of 704 bp in length  
 \* 8706 8805: gap of 100 bp  
 \* 8806 9503: contig of 698 bp in length  
 \* 9504 9603: gap of 100 bp  
 \* 9604 10289: contig of 686 bp in length  
 \* 10290 10389: gap of 100 bp  
 \* 10390 11101: contig of 712 bp in length  
 \* 11102 11201: gap of 100 bp  
 \* 11202 11908: contig of 707 bp in length  
 \* 11909 12008: gap of 100 bp  
 \* 12009 12714: contig of 706 bp in length

\* 12715 12814: gap of 100 bp  
 \* 12815 13516: contig of 702 bp in length  
 \* 13517 13616: gap of 100 bp  
 \* 13617 14321: contig of 705 bp in length  
 \* 14322 14421: gap of 100 bp  
 \* 14422 15117: contig of 696 bp in length  
 \* 15118 15217: gap of 100 bp  
 \* 15218 15922: contig of 705 bp in length  
 \* 15923 16022: gap of 100 bp  
 \* 16023 16734: contig of 712 bp in length  
 \* 16735 16834: gap of 100 bp  
 \* 16835 17543: contig of 709 bp in length  
 \* 17544 17643: gap of 100 bp  
 \* 17644 18339: contig of 696 bp in length  
 \* 18340 18439: gap of 100 bp  
 \* 18440 19142: contig of 703 bp in length  
 \* 19143 19242: gap of 100 bp  
 \* 19243 19956: contig of 714 bp in length  
 \* 19957 20056: gap of 100 bp  
 \* 20057 20780: contig of 724 bp in length  
 \* 20781 20880: gap of 100 bp  
 \* 20881 21599: contig of 719 bp in length  
 \* 21600 21699: gap of 100 bp  
 \* 21700 22381: contig of 682 bp in length  
 \* 22382 22481: gap of 100 bp  
 \* 22482 23192: contig of 711 bp in length  
 \* 23193 23292: gap of 100 bp  
 \* 23293 23998: contig of 706 bp in length  
 \* 23999 24098: gap of 100 bp  
 \* 24099 24807: contig of 709 bp in length  
 \* 24808 24907: gap of 100 bp  
 \* 24908 25602: contig of 695 bp in length  
 \* 25603 25702: gap of 100 bp  
 \* 25703 26316: contig of 614 bp in length  
 \* 26317 26416: gap of 100 bp  
 \* 26417 27125: contig of 709 bp in length  
 \* 27126 27225: gap of 100 bp  
 \* 27226 27934: contig of 709 bp in length  
 \* 27935 28034: gap of 100 bp  
 \* 28035 28757: contig of 723 bp in length  
 \* 28758 28857: gap of 100 bp  
 \* 28858 29554: contig of 697 bp in length  
 \* 29555 29654: gap of 100 bp  
 \* 29655 30358: contig of 704 bp in length  
 \* 30359 30458: gap of 100 bp  
 \* 30459 31165: contig of 707 bp in length  
 \* 31166 31265: gap of 100 bp  
 \* 31266 31970: contig of 705 bp in length  
 \* 31971 32070: gap of 100 bp  
 \* 32071 32770: contig of 700 bp in length  
 \* 32771 32870: gap of 100 bp  
 \* 32871 33578: contig of 708 bp in length  
 \* 33579 33678: gap of 100 bp  
 \* 33679 34390: contig of 712 bp in length  
 \* 34391 34490: gap of 100 bp  
 \* 34491 35199: contig of 709 bp in length  
 \* 35200 35299: gap of 100 bp  
 \* 35300 35998: contig of 699 bp in length  
 \* 35999 36098: gap of 100 bp  
 \* 36099 36781: contig of 683 bp in length  
 \* 36782 36881: gap of 100 bp  
 \* 36882 37601: contig of 720 bp in length  
 \* 37602 37701: gap of 100 bp  
 \* 37702 38409: contig of 708 bp in length  
 \* 38410 38509: gap of 100 bp  
 \* 38510 39210: contig of 701 bp in length  
 \* 39211 39310: gap of 100 bp  
 \* 39311 40022: contig of 712 bp in length  
 \* 40023 40122: gap of 100 bp  
 \* 40123 40834: contig of 712 bp in length  
 \* 40835 40934: gap of 100 bp  
 \* 40935 41660: contig of 726 bp in length  
 \* 41661 41760: gap of 100 bp



Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,  
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwotu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,  
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,X., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Unpublished  
2 (bases 1 to 152895)  
Worley,K.C.  
Direct Submission  
Submitted (30-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 152895)  
Worley,K.C.  
Direct Submission  
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20340380.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GVPZ  
Center clone name: CH230-137H14  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 91714 bases at least Q40  
Consensus quality: 98583 bases at least Q30  
Consensus quality: 103878 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 54 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1041: contig of 1041 bp in length  
\* 1042 1141: gap of unknown length  
\* 1142 2262: contig of 1121 bp in length  
\* 2263 2362: gap of unknown length  
\* 2363 3497: contig of 1135 bp in length  
\* 3498 3597: gap of unknown length  
\* 3598 5210: contig of 1613 bp in length  
\* 5211 5310: gap of unknown length  
\* 5311 6936: contig of 1626 bp in length  
\* 6937 7036: gap of unknown length  
\* 7037 8291: contig of 1255 bp in length  
\* 8292 8391: gap of unknown length  
\* 8392 9663: contig of 1272 bp in length  
\* 9664 9763: gap of unknown length  
\* 9764 11369: contig of 1606 bp in length  
\* 11370 11469: gap of unknown length  
\* 11470 13073: contig of 1604 bp in length  
\* 13074 13173: gap of unknown length  
\* 13174 14521: contig of 1348 bp in length  
\* 14522 14621: gap of unknown length  
\* 14622 16274: contig of 1653 bp in length  
\* 16275 16374: gap of unknown length  
\* 16375 17405: contig of 1031 bp in length  
\* 17406 17505: gap of unknown length  
\* 17506 18547: contig of 1042 bp in length  
\* 18548 18647: gap of unknown length  
\* 18648 19908: contig of 1261 bp in length  
\* 19909 20008: gap of unknown length  
\* 20009 21460: contig of 1452 bp in length  
\* 21461 22822: contig of 1262 bp in length  
\* 22823 22922: gap of unknown length  
\* 22923 24438: contig of 1515 bp in length  
\* 24439 25832: contig of 1295 bp in length  
\* 25833 25932: gap of unknown length  
\* 25933 27104: contig of 1172 bp in length  
\* 27105 27204: gap of unknown length  
\* 27205 29052: contig of 1848 bp in length  
\* 29053 30398: contig of 1246 bp in length  
\* 30399 32426: contig of 1928 bp in length  
\* 32427 32526: gap of unknown length  
\* 32527 34955: contig of 2429 bp in length  
\* 34956 35055: gap of unknown length  
\* 35056 36642: contig of 1587 bp in length  
\* 36643 36742: gap of unknown length  
\* 36743 37820: contig of 1078 bp in length  
\* 37821 37920: gap of unknown length  
\* 37921 39394: contig of 1474 bp in length  
\* 39395 39494: gap of unknown length  
\* 39495 40928: contig of 1434 bp in length  
\* 40929 41028: gap of unknown length  
\* 41029 43380: contig of 2352 bp in length  
\* 43381 43480: gap of unknown length  
\* 43481 44748: contig of 1268 bp in length  
\* 44749 44848: gap of unknown length  
\* 44849 46927: contig of 2079 bp in length  
\* 46928 47027: gap of unknown length  
\* 47028 49581: contig of 2454 bp in length  
\* 49582 49581: gap of unknown length  
\* 49582 51923: contig of 2342 bp in length  
\* 51924 52023: gap of unknown length  
\* 52024 54126: contig of 2103 bp in length  
\* 54127 54226: gap of unknown length  
\* 54227 57214: contig of 2988 bp in length  
\* 57215 57314: gap of unknown length  
\* 57315 60024: contig of 2710 bp in length  
\* 60025 63163: contig of 3039 bp in length  
\* 63164 63263: gap of unknown length  
\* 63264 66703: contig of 3440 bp in length  
\* 66704 66803: gap of unknown length  
\* 66804 68557: contig of 1754 bp in length  
\* 68558 72882: contig of 4225 bp in length  
\* 72883 72982: gap of unknown length  
\* 72983 76542: contig of 3560 bp in length



PUBMED	2246248
FEATURES	
source	
Location/Qualifiers	
1..3327	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/map="19p13.1"	
/clone="pHCMF2"	
1..3327	
/gene="CRTM"	
<1..1023	
/gene="CRTM"	
/codon_start=1	
/evidence-experimental	
/product="cartilage matrix protein"	
/protein_id="AAAG3904.1"	
/db_xref="GI:180654"	
/db_xref="GDB:G00-127-280"	
/translation="PODSVODVSARASGVELFATGVSDKATLRQIASPEQDEHY DYVESYVIEKLKRKFQACVVDLCATGDHCEQCISSPSGYTCACHEGTFLNSD KYNVCSSGGGSSATDLVELIDGSXSRPENLELVKKFIQSIVDTLDVSKLQAQGL GYVSSSVRKEPPLGRFTDKDKAAVRNMSMEKMTGAALKYLILDNSFTVSSGARF GOAQVGIVFTDGRSQDIINDAKKAKDLGFKFAVGVGNAVEDELREIASEPVAEHYF YTADPKTINQIGKLLQKKICVEEDPCACESLVKFKQAVEGLLAQLRKLEAVSKRLAI LENTVV"	
<1..196	
/gene="CRTM"	
/note="G00-127-280"	
/number=3	
197..322	
/gene="CRTM"	
/note="G00-127-280"	
/number=4	
323..739	
/gene="CRTM"	
/note="G00-127-280"	
/number=5	
740..892	
/gene="CRTM"	
/note="G00-127-280"	
/number=6	
893..973	
/gene="CRTM"	
/note="G00-127-280"	
/number=7	
974..>3327	
/gene="CRTM"	
/note="G00-127-280"	
/number=8	
BASE COUNT	769 a 823 c 949 g 786 t
ORIGIN	
Query Match	2.8%; Score 95.6; DB 9; Length 3327;
Best Local Similarity	50.4%; Pred. No. 8.3e-10;
Matches	264; Conservative 0; Mismatches 254; Indels 6; Gaps 1;
QY	422 AAACCATCGGGAAGATTTCAGCTGCCACAAAATGATGTGCTCGCCTGCAGTGGAACA 481
Db	299 ACAGCGACGCAAGACTGCAATGTCTCGAGTTGGTGGTGGCAGCTCGGCCACTGACC 358
QY	482 TCATGTTTTCTGTAGATGGTCTAACAGCGTCGGGAAGGAGGACTTTGAAGAAGTCCAAGC 541
Db	359 TGGTCTTCTCATTCAGCGGATCCAAAGTGTGAGGCCAGAACCTTGAAGTGGTGAAGA 418
QY	542 ACTTTGCCATCACACTCTGACGCTCTGGACATCAGCCCCGAGGAGGCTCAGATGGGAG 601
Db	419 AGTTCATCATCGATCGTGGATACGCTGGACGCTGCAGAACGCTGGCCCAGGTGGGGC 478
QY	602 CATTCAGTTTCAGTCCCACTCTCATCTCGGAATTCCTCCCTGGATTCAATTTCAACCCCAAC 661
Db	-479 TGGTGCAGTACTCAAGCTCTGTGCCCGCAGGAGTTCCTCCCTGGTCTGCTCCACACCAAGA 538
QY	662 AGGAAGTAGGCAAGAATAAAGAGAGTGGTTTTCAAAGGAGGCGCGCAGGAGCGGAAC 721

PGDAVLAWAAAPSGVLVYQITWTPLGEGKAHEISVPNGLGTAVLPGIGRHHTEYDVT  
ILAYRDGARDSPVLSURTYTPTSVRGPPNNALASVETPSDLSQVSTWPLGRLVHYLMT  
YAPAGLGPKEKSPVSCAKSHVTLPDLOAATKYRVLVSAIYAAGRSSEVASATGQATCP  
ALRPDGLSPGFDLWAFSAIVEKAYASIRGVAMEPSPAFGGTPTFLPKDAQLTFRVSDV  
YAPLPPEHTIVFLVRLPETHPREAFALWQMTAEDFQPLLVGLLDAGKAKSLTVFHRDP  
RAALQQLDPOEVRKIFGTSFHKHVAHVSRVLYVDCRKRAEPLBGMGSPPAAG  
FVTLGRALKARPRSSAAFLQMLQIVGSDTWADRCCELPSADRGTCFPAFVSC  
SCSSTPPGPGOGPCLPGRNCTGEOGFPGPGRPPGVYKEGKDHLGLPCLQGLPGHGO  
GIPGRVLOGPGRMGRLGCTAGLPGPPRPGFGOMAGARGTSGERGPPVTGVTGLPG  
PKGERKEGPEQSLATLYLVSQACSAIQTHVSKFSDSFHENTRPMPLILEQKLEPGT  
EPUGSPGTSKALPVGEWGRGHRHCEGPEPAGQMGSPGOQGAQSTQGLWE".

1530 a 2336 c 2283 g 1355 t

BASE COUNT  
ORIGIN

<b>Query Match</b>	<b>2.6%</b>	<b>Score 87.8;</b>	<b>DB 9;</b>	<b>Length 7504;</b>
<b>Best Local Similarity</b>	<b>51.0%;</b>	<b>Pred. No. 3.6e-08;</b>		
<b>Matches 237; Conservative</b>	<b>0;</b>	<b>Mismatches 22;</b>	<b>Indels 6;</b>	<b>Gaps 1;</b>

  

QY	473	CAGTGGACATCATGTTTCCTGTAGATGGTGCTACACAGCGTCGGGAAGGGAGCCTTTGAAA	532
Db			
Db	82	CTGCTGACATTGGTCTCTGGTGGACGGTCTCTGGACATTTGGCCACAGTCACACTTCACGC	141
QY	533	GGTCCAAGCACTTTGCCATCACAGCTGTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCA	592
Db			
Db	142	AGGTCAAAGGAACTTCCTGGCCAGTGTCACTGCACCCTTTGAAATCGGCCGCGATTAAGTGCC	201
QY	593	GAGTGGGAGCATTCACAGTTCAGTTTTCCACTCTCATCTGGAATTCCTCTTGATTCATTTT	652
Db			
Db	202	AAGTAGCCCTGCATCACTACAGCGGGGATGCTCAGACTGAGTGGGACCTGNACTCCCCTCA	261
QY	653	CAACCCAACAGGAAGTGAAGGCAAGAATAAAGAGGATGGTTTCCAAAGAGGGCGCACGG	712
Db			
Db	262	GCACAAGGAAACAGTGTCTGCAGCTGTGGCGCCCTCCGCTACAAAGGGGGGAAACAGT	321
QY	713	AGACGGAACTTGCTCTCAAAATACCTTCT-----GCACAGAGGGTTCCTTGAGGCAGAA	766
Db			
Db	322	TCACAGGCCCTTGCCCTGACCCACGTGCTGGGGCAGAACCTGCACCGCGCTGGCCTCC	381
QY	767	ATGCTTCTGTGCCCCAGATCCTCATCATCTGCTACTGATGGGAAGTCCACAGGGGATGTGG	826
Db			
Db	382	GTCACAGAGGCAGCCAAGGTGGTGATTCTGTGTGACGGACGGCAAGTCCACAGACCATGTGC	441
QY	827	CAGTGGCATCCAAGCAGCTGAAGAAAAAGGGGTGTCACTGTGTTTTGCTGTGGGGTCAAGT	886
Db			
Db	442	ACACTGTGCCCCGTGCTCTCNAAGGACCTGGCGGTGAACGCTTCGCTGTGGGTGTGAAGA	501
QY	887	TTOCCAGGTGGGAGGAGCTGCATGCACTGCGCCACGCGAGCCCTAGAG	931
Db			
Db	502	ACGCCGATGAGGCTGAGCTGAGGCTCTCTGCGCTCCCGCCGCGAGG	546

RESULT 13					
AC112730/c					
LOCUS	AC112730	19413 bp	DNA	linear	HTG 10-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-8B11, *** SEQUENCING IN PROGRESS ***, 70 unordered pieces.				

AC112730.2	GI:21716976
VERSION	HTG; HTGS_PHASE1.
KEYWORDS	Norway rat.
SOURCE	<i>Rattus norvegicus</i>
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 194143)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola, B., Ali-osman,F.R., Allen,C., Allbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbia,J., Benton,J., Bimake,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
Homsfi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,  
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louiseged, H.,  
Lozaro, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Mosser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwu, S., Oguh, M., Okuwonu, G.,  
Oreguene, N., Ovielo, R., Pace, A., Payton, B., Perry, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojokhan, I., Roife, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Verd, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, J.,  
Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 194143)  
Worley, K.C.

Direct Submission  
Submitted (24-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 194143)  
Worley, K.C.

Direct Submission  
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jul 9, 2002 this sequence version replaced gi:18874114.

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: GCPA  
Center Clone name: CH230-Ball1  
----- Summary Statistics -----  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.930329  
Consensus quality: 131293 bases at least Q40  
Consensus quality: 137579 bases at least Q30  
Consensus quality: 142936 bases at least Q20

-----

\*\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\*\* NOTE: This is a 'working draft' sequence. It currently  
consists of 70 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.





```
/chromosome="2"
/clone="RP23-462016"
/clone_lib="RPCI-23"
misc_feature 104714..104881
    /note="Sequence from uni-directional dGTP big dye
    terminator reads only"
misc_feature 140443..140460
    /note="Single clone region. Sequence from reads from a
    short insert library derived from a clone PCR.
    Restriction digest data confirm the assembly."
misc_feature 140460
    /note="Tandem repeat. Forced join. Gap size estimated to
    be approximately 900bp by restriction digest data."
BASE COUNT 55113 a 43425 c 42104 g 50027 t
ORIGIN
Query Match 2.4%; Score 82; DB 10; Length 190669;
Best Local Similarity 52.7%; Pred. No. 5.3e-07;
Matches 178; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 1893 GCCAGGTCGCCGACACAAAGCCCTGGACCTCGTCTTCATGTTGGACACCTCTGCCTCAGT 1952
    |||||
Db 135490 GCCAGGTCGCCGAGGGCCACGTGGATCTTCTCCTGGTGGATGTTCCAGAGCCGT 135431
    |||||
QY 1953 AGGCGCCGAGATTTCCTCAGATGACAGCTTTGTGAGAAAGCTGTGCCCTCCAGTTTGA 2012
    |||||
Db 135430 GCGCCACAGAACTTCGAGCTGCTGAAGCGCTTCGTGAACCCAGATTCTGGATTCCTTGA 135371
    |||||
QY 2013 GGTGAACCTGACGTGACACAGCTCGCCCTGGTGGTGTATGGCAGCCAGTGCAGACTGC 2072
    |||||
Db 135370 CGTCTCTCCGAGGGCACACGTGTGGCTGGTGCAAGTTCTCCAGCGGGTGGCACCCGA 135311
    |||||
QY 2073 CTTCCGGCTGGACACCAAAACCCACCGGGCTGCGATGCTCGGGGCCATTAGCCAGGCCCC 2132
    |||||
Db 135310 GTTCCCTCGGCCGCTATGGCACCCGACGCTGAGTGAACCCAGCAGTTTGGCCCTGGA 135251
    |||||
QY 2133 CTACCTAGTGGGGTGGGCTACCGGCCACCGCCCTGCTGCACATCTATGACAAAGTGAT 2192
    |||||
Db 135250 GTACATGGAGCGGGCACCAATGACAGGGCTGGCCCTGCGTGCATGCTGGAGCACAGCTT 135191
    |||||
QY 2193 GACCGTCCAGAGGGGTGCCCGGCTGGTGTCGCCCAAG 2230
    |||||
Db 135190 CTCTAGGGCCAGGGCGCGGCTCGGACCTCAACG 135153
    |||||
```

Search completed: May 5, 2003, 04:52:15  
Job time : 8710 secs

**THIS PAGE BLANK (USPTO)**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

XX WPI; 2002-471335/50.

XX P-PSDB; ABG61891.

XX Detecting a prostate cancer-associated transcript in a cell in a  
PT patient, useful for diagnosing prostate cancer (PC) or screening  
PT modulators of PC, by determining if prostate cancer-associated genes  
PT are expressed in a prostate tissue

XX Claim 22; Page 376-377; 436pp; English.

XX The present invention relates to methods of detecting a prostate  
CC cancer-associated transcript in a cell from a patient. The method  
CC comprises contacting a biological sample from the patient with  
CC prostate cancer-associated polynucleotides (designated PC genes) that  
CC selectively hybridize to a sequence that is at least 80% identical  
CC to them. The prostate cancer-associated polynucleotide sequences  
CC are differentially expressed in prostate tumour tissue or in  
CC prostate cancer and are derived from the tissues of various  
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
CC The methods of the invention are useful for diagnosing and treating  
CC prostate cancer in mammals. The prostate cancer-associated genes are  
CC useful for diagnosing or treating prostate cancer, as well as for  
CC identifying modulators of prostate cancer or agents that inhibit  
CC prostate cancer. The nucleic acid sequences are particularly useful  
CC in gene therapy, as a vaccine or in antisense applications.  
CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide  
CC sequences.

XX Sequence 3375 BP; 660 A; 944 C; 1062 G; 709 T; 0 other;

Query Match 100.0%; Score 3375; DB 24; Length 3375;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGTGTTCGGGGTGCACCGTCCGAGCTGGGTGGTGACCCCGGTAGAAGTACTT 60

DB 1 GACAGTGTTCGGGGTGCACCGTCCGAGCTGGGTGGTGACCCCGGTAGAAGTACTT 60

QY 61 TTTTATTTCAGACCTGGCGGATCCGCTTTAAACCGCAGGGGCTATGACCTC 120

DB 61 TTTTATTTCAGACCTGGCGGATCCGCTTTAAACCGCAGGGGCTATGACCTC 120

QY 121 CTGCGGGTAGTTCCTCCGACCTACCGGGTGGGTGGTGGCGCTCTCCAGGAGAG 180

DB 121 CTGCGGGTAGTTCCTCCGACCTACCGGGTGGGTGGTGGCGCTCTCCAGGAGAG 180

QY 181 ACAACAGGTGTCACAGTGGAGCGCGCGCGCGCGCGCTCTGTGATCCCGTAGCGG 240

DB 181 ACAACAGGTGTCACAGTGGAGCGCGCGCGCGCGCGCTCTGTGATCCCGTAGCGG 240

QY 241 CCCCCTGGCCGAGCGCGCGCGGGTCTGTGATGACGCGCGCGCGCGCGCGCTGG 300

DB 241 CCCCCTGGCCGAGCGCGCGCGGGTCTGTGATGACGCGCGCGCGCGCGCTGG 300

QY 301 TGCGCGCTCTCCTTCGGTTATATCAACATGCCCGCTTTCTGTGTTGAGGCGGCTGT 360

DB 301 TGCGCGCTCTCCTTCGGTTATATCAACATGCCCGCTTTCTGTGTTGAGGCGGCTGT 360

QY 361 GTTTTCTCTTTTCCAGAGTCCCGCTCTCTCCCTCTCCAGGAAGTCCATGTAACAAA 420

DB 361 GTTTTCTCTTTTCCAGAGTCCCGCTCTCTCCCTCTCCAGGAAGTCCATGTAACAAA 420

QY 421 GAAACCATCGGGAAGATTTTCAGTGCACGACAAATGATGTGGTCTCGCTGAGTGGAC 480

DB 421 GAAACCATCGGGAAGATTTTCAGTGCACGACAAATGATGTGGTCTCGCTGAGTGGAC 480

QY 481 ATCATGTTCTGTAGTGGTCTAACAGCGTGGGAAAGGAGCTTTTGAAGGTTCCAAG 540

DB 481 ATCATGTTCTGTAGTGGTCTAACAGCGTGGGAAAGGAGCTTTTGAAGGTTCCAAG 540

QY 541 CACTTTGCCATCACATCTGTGTGACGGTCTGTGGACATCAGCCCCCAGAGGGTCTCAGAGTGGGA 600  
DB 541 CACTTTGCCATCACATCTGTGTGACGGTCTGTGGACATCAGCCCCCAGAGGGTCTCAGAGTGGGA 600  
QY 601 GCATTCCAGTTCAGTTCCTCTCTCATCTGGAATTCCTCTGATTCATTTTCAACCCAA 660  
DB 601 GCATTCCAGTTCAGTTCCTCTCTCATCTGGAATTCCTCTGATTCATTTTCAACCCAA 660  
QY 661 CAGGAAGTGAAGCAAGAATCAAGAGGATGGTGTTCAAAGAGGGGCGCAGAGAGCGAA 720  
DB 661 CAGGAAGTGAAGCAAGAATCAAGAGGATGGTGTTCAAAGAGGGGCGCAGAGAGCGAA 720  
QY 721 CTTGCTCTGAAATACCTTCTGCACAGAGGTTGCCTGGAGGCAGAAATGCTTCTGTGCC 780  
DB 721 CTTGCTCTGAAATACCTTCTGCACAGAGGTTGCCTGGAGGCAGAAATGCTTCTGTGCC 780  
QY 781 CAGATCCTCATCATCTCTCATCTGATGGAAGTCCAGGGGGATGTGCACATGCCATCCAAG 840  
DB 781 CAGATCCTCATCATCTCTCATCTGATGGAAGTCCAGGGGGATGTGCACATGCCATCCAAG 840  
QY 841 CAGCTGAAGGAAAGGGTGTCTACTGTGTGTGGGGTCAAGTTTCCAGGTGGGAG 900  
DB 841 CAGCTGAAGGAAAGGGTGTCTACTGTGTGTGGGGTCAAGTTTCCAGGTGGGAG 900  
QY 901 GAGCTGCATGCATCTGCCAGCGAGCTTAGAGGCAGACGCTGTGTGGTGTGACAGGTG 960  
DB 901 GAGCTGCATGCATCTGCCAGCGAGCTTAGAGGCAGACGCTGTGTGGTGTGACAGGTG 960  
QY 961 GAGGATGCCACCAACCGCTCTTACAGCAGCTTCAGCAGCTCGGCCATCTCTCCAGGCC 1020  
DB 961 GAGGATGCCACCAACCGCTCTTACAGCAGCTTCAGCAGCTCGGCCATCTCTCCAGGCC 1020  
QY 1021 ACSCCAGACTGCAGGGTTCAGGGTCAACCCCTGTGAGCAGACAGCGTGGAGATGGTCCGG 1080  
DB 1021 ACSCCAGACTGCAGGGTTCAGGGTCAACCCCTGTGAGCAGACAGCGTGGAGATGGTCCGG 1080  
QY 1081 GAGTTCGCTGGCAATGCCCATCTCTGAGAGGATTCGCGGGGACCCCTTGGCGGTCTGGCT 1140  
DB 1081 GAGTTCGCTGGCAATGCCCATCTCTGAGAGGATTCGCGGGGACCCCTTGGCGGTCTGGCT 1140  
QY 1141 GCACACTGTCCCTTCTACAGCTGGAAGAGAGTCTTCTTACCCACCTGCGACCTGTCTAC 1200  
DB 1141 GCACACTGTCCCTTCTACAGCTGGAAGAGAGTCTTCTTACCCACCTGCGACCTGTCTAC 1200  
QY 1201 AGGACACCTGCGCCAGCGCCCTGTGACTCGAGCCCTGCGAGAAATGGAGACATGTGTT 1260  
DB 1201 AGGACACCTGCGCCAGCGCCCTGTGACTCGAGCCCTGCGAGAAATGGAGACATGTGTT 1260  
QY 1261 CCAGAGGAGTGCAGCGCTACCACTGCTCTGCGCGCTGGCCCTTGGAGGGAGGCTAAC 1320  
DB 1261 CCAGAGGAGTGCAGCGCTACCACTGCTCTGCGCGCTGGCCCTTGGAGGGAGGCTAAC 1320  
QY 1321 TGTGCCCTGAAGCTGAGCCTGGAATGACGGTTCAGCTCTCTCTCTCTCTGCTGGACAGCTCT 1380  
DB 1321 TGTGCCCTGAAGCTGAGCCTGGAATGACGGTTCAGCTCTCTCTCTCTCTGCTGGACAGCTCT 1380  
QY 1381 GCGGGCACCACCTCTGACAGCGCTTCTCGCGGGCCAAAGTCTTCTGTAAGCGGTTTGGCGG 1440  
DB 1381 GCGGGCACCACCTCTGACAGCGCTTCTCGCGGGCCAAAGTCTTCTGTAAGCGGTTTGGCGG 1440  
QY 1441 GCGGCTCTGAGCGAGCAGCTCTCGGGCCGAGTGGGTGTGGCCACATACAGAGGAGCTG 1500  
DB 1441 GCGGCTCTGAGCGAGCAGCTCTCGGGCCGAGTGGGTGTGGCCACATACAGAGGAGCTG 1500  
QY 1501 CTGGTGGCGGTGCTCTGGGGAGTACCAGGATGTGCCTGACCTGTCTTGAGGCTCGAT 1560  
DB 1501 CTGGTGGCGGTGCTCTGGGGAGTACCAGGATGTGCCTGACCTGTCTTGAGGCTCGAT 1560  
QY 1561 GGCATTCCTTCCGTGGTGGCGCCACCTGTACGGGCGAGTGCCTTGGCGGAGCGGGCAGAG 1620  
DB 1561 GGCATTCCTTCCGTGGTGGCGCCACCTGTACGGGCGAGTGCCTTGGCGGAGCGGGCAGAG 1620

QY 1621 CGTGGCTTCGGGAGCGCCACAGACAGGCGCAGGACCGGCGCACCTAGAGTGGTGGTTTG 1680  
Db 1621 CGTGGCTTCGGGAGCGCCACAGACAGGCGCAGGACCGGCGCACCTAGAGTGGTGGTTTG 1680  
QY 1681 CTCAGTGAAGTACACTCCGAGGATGAGTTCGCGGCGCCAGCGCTCAGCAAGGCGCGA 1740  
Db 1681 CTCAGTGAAGTACACTCCGAGGATGAGTTCGCGGCGCCAGCGCTCAGCAAGGCGCGA 1740  
QY 1741 GAGTGTCTTCCTGCTGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 1800  
Db 1741 GAGTGTCTTCCTGCTGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 1800  
QY 1801 GGCAGCCCAAGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 1860  
Db 1801 GGCAGCCCAAGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 1860  
QY 1861 GAGCTGAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 1920  
Db 1861 GAGCTGAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 1920  
QY 1921 CTCGCTTCATGTTGAGACCTCTGCTCAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 1980  
Db 1921 CTCGCTTCATGTTGAGACCTCTGCTCAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 1980  
QY 1981 AGCTTGTGAGAGCTGTGCTCAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 2040  
Db 1981 AGCTTGTGAGAGCTGTGCTCAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 2040  
QY 2041 CTGCTGTGATGACGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2100  
Db 2041 CTGCTGTGATGACGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2100  
QY 2101 GCTGCGATGCTGGGCGCATAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2160  
Db 2101 GCTGCGATGCTGGGCGCATAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2160  
QY 2161 ACCGCTGCTGACATCTATGACAAAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2220  
Db 2161 ACCGCTGCTGACATCTATGACAAAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2220  
QY 2221 GTCCCCAAAGCTGTGCTGCTCAGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2280  
Db 2221 GTCCCCAAAGCTGTGCTGCTCAGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2280  
QY 2281 GCCCAGAGCTGAGGAGCAATGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 2340  
Db 2281 GCCCAGAGCTGAGGAGCAATGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 2340  
QY 2341 AGTGAGGCTGTGCGGAGGCTGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2400  
Db 2341 AGTGAGGCTGTGCGGAGGCTGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2400  
QY 2401 GCCGACCTGTGCGGAGGCTGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2460  
Db 2401 GCCGACCTGTGCGGAGGCTGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2460  
QY 2461 CCAGTCAACCTGTGCAAAACCCAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2520  
Db 2461 CCAGTCAACCTGTGCAAAACCCAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2520  
QY 2521 GGGAGCTACCGCTGCAAGTGTGCGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2580  
Db 2521 GGGAGCTACCGCTGCAAGTGTGCGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2580  
QY 2581 TGGAGCTCTTGTCTGTATGTGAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2640  
Db 2581 TGGAGCTCTTGTCTGTATGTGAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAG 2640  
QY 2641 ATGGCTCCCTGTGAGGAGGAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2700  
Db 2641 ATGGCTCCCTGTGAGGAGGAGGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 2700  
QY 2701 GGCAGTGAATGGTGCCTACCTTCTGGAATGTCTGTGCGCCCGGAGTCTTGAATGTCTGC 2760

Db 2701 GGCAGTGAATGGTGCCTACCTTCTGGAATGTCTGTGCGCCCGGAGTCTTGAATGTCTGC 2760  
QY 2761 TTCCGCGCGTGGCCAGGACCACTATTCTCTACTGAGGAGGAGGAGTGTCCCAACTGCGAGCC 2820  
Db 2761 TTCCGCGCGTGGCCAGGACCACTATTCTCTACTGAGGAGGAGGAGTGTCCCAACTGCGAGCC 2820  
QY 2821 ATGCTGCTTAGAGACAAGAAAGAGCTGTATCTACCCACAAACGATGTGTGTTGAAAAGTT 2880  
Db 2821 ATGCTGCTTAGAGACAAGAAAGAGCTGTATCTACCCACAAACGATGTGTGTTGAAAAGTT 2880  
QY 2881 TTGATGTGAAGTAATACCCACTTTCTGTACCTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
Db 2881 TTGATGTGAAGTAATACCCACTTTCTGTACCTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
QY 2941 CTGCGACCTTTCCCTTCAGGATAAACAAAGGGTCTCTGAAGACTTTAAATTTAGCGGCTCTGA 3000  
Db 2941 CTGCGACCTTTCCCTTCAGGATAAACAAAGGGTCTCTGAAGACTTTAAATTTAGCGGCTCTGA 3000  
QY 3001 CGTTCTTTGCACAAATCAATGCTGCCAGAAATGTTGTTGACACAGTAATGCTGCGGCTCTGA 3060  
Db 3001 CGTTCTTTGCACAAATCAATGCTGCCAGAAATGTTGTTGACACAGTAATGCTGCGGCTCTGA 3060  
QY 3061 AGGCTTTACTAGAGCATCTTTGGAGCGGCGAAGGCCACGCGCTTTCAAGATGGAAGCA 3120  
Db 3061 AGGCTTTACTAGAGCATCTTTGGAGCGGCGAAGGCCACGCGCTTTCAAGATGGAAGCA 3120  
QY 3121 GCAGCTTTCCACTTCCCGAGAGACATCTCGGATGCAATTTGCATTTGAGTCTGAAAGGGG 3180  
Db 3121 GCAGCTTTCCACTTCCCGAGAGACATCTCGGATGCAATTTGCATTTGAGTCTGAAAGGGG 3180  
QY 3181 CTGAGGAGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240  
Db 3181 CTGAGGAGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240  
QY 3241 GGTCTCAGACTGAATGTGACCAATTAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300  
Db 3241 GGTCTCAGACTGAATGTGACCAATTAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300  
QY 3301 TGTGATGCGGCGGAGTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360  
Db 3301 TGTGATGCGGCGGAGTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360  
QY 3361 ACCTTGAAGTCTTC 3375  
Db 3361 ACCTTGAAGTCTTC 3375

## RESULT 2

AAS81530

ID AAS81530 standard; cDNA; 3485 BP.

XX AAS81530;

AC AC

XX XX

DT 13-FEB-2002 (first entry)

XX XX

DE DNA encoding novel human diagnostic protein #17334.

XX XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; as.

XX XX

OS Homo sapiens.

XX XX

PN WO200175067-A2.

XX XX

PD 11-OCT-2001.

XX XX

PF 30-MAR-2001; 2001WO-US08631.

XX XX

PR 31-MAR-2000; 2000US-0540217.

XX XX

PR 23-AUG-2000; 2000US-0649167.

XX XX

PA (HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;  
WPI; 2001-639362/73.  
P-PSDB; ABG17343.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 1; SEQ ID No 17334; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (III). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

SEQ ID No 17334; 103pp; English.

Query Match 59.9%; Score 2020.8; DB 23; Length 3485;  
Best Local Similarity 92.9%; Pred. No. 0;  
Matches 2185; Conservative 0; Mismatches 32; Indels 135; Gaps 1

400 CAGGAAGTCCATGTAAGCAAGAAACCATCGGGAAGATTTCAGCTGCCACGCAAAATGATG 459  
|||||  
1269 CAGGAAGTCCATGTAAGCAAGAAACCATCGGGAAGATTTCAGCTGCCACGCAAAATGATG 1328  
|||||

460 TGTGCTCGGCTGCAGTGGACATCATGTTCTGTGTAGATGGGTCTAACACGCTCGGAAA 519  
|||||

1329 TGTGCTCGGCTGCAGTGGACATCATGTTCTGTGTAGATGGGTCTAACACGCTCGGAAA 1388  
|||||

520 GGGAGCTTTGAAGGTCCAAGCACATTTGCCATCACAGTCTGTGACGCTCTGGACATCAGC 579  
|||||

1389 GGGAGCTTTGAAGGTCCAAGCACATTTGCCATCACAGTCTGTGACGCTCTGGACATCAGC 1448  
|||||

580 CCCGAGAGGTCAGAGTGGAGCATTCACAGTTCAGTTCCTCCTCATCTCGGAATTCCTCC 639  
|||||

1449 CCCGAGAGGTCAGAGTGGAGCATTCACAGTTCAGTTCCTCCTCATCTCGGAATTCCTCC 1508  
|||||

640 TTGGATTATTTCAACCCACAGGAGTGAAGGCAAGAATCAAGAGGATGGTTTTCAAA 699  
|||||

1509 TTGGATTATTTCAACCCACAGGAGTGAAGGCAAGAATCAAGAGGATGGTTTTCAAA 1568  
|||||

700 GGAGGCGGCACGAGACGGAATTCCTCGAAATACCTTTCGCACAGAGGTTGCCTGGA 759  
|||||

1569 GGAGGCGGCACGAGACGGAATTCCTCGAAATACCTTTCGCACAGAGGTTGCCTGGA 1628  
|||||

760 GGCAGAAATGCTTCTGTGCCCCAGATCCTCATCTATCGTCACTGATGGAAAGTCCACAGGG 819  
|||||

1629 GGCAGAAATGCTTCTGTGCCCCAGATCCTCATCTATCGTCACTGATGGAAAGTCCACAGGG 1688  
|||||

820 GATGTGGCACTGCCATCCCAAGCAGCTGAAGGAAAGGGGTGTCAGTGTGTTGCTGTGGG 879  
|||||

1689 GATGTGGCACTGCCATCCCAAGCAGCTGAAGGAAAGGGGTGTCAGTGTGTTGCTGTGGG 1748  
|||||



QY 722 TTGCTCTGAATACCTTTCACAGAGAGGTTGCCTGGAGGCGAGAAATGCTTCTGTGCCCC 781  
DB 521 TAGCCCTGAAA---CGCCTGAGCAGAGAGGTTCCCGGAGGCGAGAAATGCTTCTGTGCCCC 577  
QY 782 AGATCTCATCATCTGCTCACTGATGGAGCTCCACAGGGGATGGACACTGCCATCAAGC 841  
DB 578 AGATTCTTATCFCTGACGAGTGGCAAGTCCACAGGGGCCCGTGGCTCTCCCGGCTAAGC 637  
QY 842 AGCTGAAGAAAGGGGTGCTACTGTGTTTGTCTGTGGGGTCAAGTTTCCACAGTGGGAGG 901  
DB 638 AGCTGAGAGAAAGGGGCATCGTCGTGTTTGGCCGTAGGAGTCCGTTTCCACAGTGGGAGC 697  
QY 902 AGCTGATGCATGCGCCACGACCTAGAGGGAGACAGTGTCTGTGGTGTGAGCAGGTGG 961  
DB 698 AGCTGCTCAGCTGGCCAGTGAAGCAAGGACCGGATGTGTGTTGGTGTGACGAAGTG 757  
QY 962 AGGATGCCACCAAGCGCTTCTCAGCACCTCAGCAGCTCGGCCATCTCTCCAGGGCCA 1021  
DB 758 AGGATGCCACCAATGGCTTCTCAGCACCTCAGCAGCTCGGCCATCTCTCAGCACCTGCTG 817  
QY 1022 CGCCAGACTGCAGGGTCGAGGCTCACCCCTGTGACACAGGACGCTGGAGATGGTCCGGG 1081  
DB 818 ATCCAGACTGCAGGGTGAACCTCATCCCTGTGAGCGGAGGACGCTGGAGACCGTCAGG 877  
QY 1082 AGTTCCCTGGCAATGCCCATGCTGAGAGAGATCGCGGGGACCCCTTGGCGTGTGGCTG 1141  
DB 878 AGCTCGCTGGCAATGCTTGTGTGTGAGAGAGATCAAGGCAAGCAGACACTGTGCTGGCTC 937  
QY 1142 CACACTGTCCCTTCTACAGTGAAGAGAGTGTTCCTAACCCACCTGCCACCTGCTGCTACA 1201  
DB 938 TGCCCTGTCCCTTCTACAGTGAAGAGAGTGTTCAGACACACCTGCCAATGCTGCTACA 997  
QY 1202 GGACCACTGCCAGGCCCTGTGACTCGACGCCCTGCCAGAAATGAGGACACATGTGTTTC 1261  
DB 998 GAACCATCTGTCCAGGCCCTGTGACTCCAGGCCCTGCCAAATGAGGACGCTGCTGCTC 1057  
QY 1262 CAGAAGGACTGGACGCTACCACTGCTGCTGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1321  
DB 1058 CAGAAGGCTGGAGTACCACTGCTGCTGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1117  
QY 1322 GTGCCCTGAAGCTGAGCTGGAATGAGGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1381  
DB 1118 GTGCCCTGAAGCTGAGCTGGAATGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1177  
QY 1382 CGGGACCACTGTGACGCGCTTCTGCGGGCCAAAGTCTTGTGAAGCGGTTTGTGCGGG 1441  
DB 1178 CAGGCACCACTATTGGGGGCTTCCGGAGGGCCAAAGGCTTTGTCAAGCGCTTTGTGCGGG 1237  
QY 1442 CGTGTGTGAGCGAGGACTCTCGGGCCCGAGTGGTGTGGCCACATACACAGGAGCTGC 1501  
DB 1238 CGTGTGTGAGGAGGACTCCCGAGCCCGGCTTGGGATAGCCAGTTATGGCAGGAATCTAA 1297  
QY 1502 TGTGTGCGGTG-CCTGTGGGGAGTACCAGGA-TGTGCTGACCTGCTGCTGAGGCTGCA 1559  
DB 1298 TGTGTGCGGTGCTGCTGCGGGAGTACAGCATTTGTGCGGACCTGATCAGGACCTTGA 1357  
QY 1560 TGGCATTCCTTCCGTGTGTGGCCCCACCTGAGCGGCACTGCTTCCGGCAGGCGCAGA 1619  
DB 1358 CAGCATTCCTTTCAGCGGTGGCCGACCCCTAACCGGAGTGCCTTCTCCAGGTGGCAGA 1417  
QY 1620 GCGTGTGCTTGGAGCGCCACAGGACAGCCAGGACCGCCACCTAGTAGTGGTGTGTTT 1679  
DB 1418 GCACGCTTGGAGTGGCCAGCAGGACTGCTCAGGACAGGCGCCAGCAGAGTAGTAGTCT 1477  
QY 1680 GCTCACTGATCACATCCGAGGATGAGTGTGGGGCCCGAGCGCTCAGCAAGGCGGG 1739  
DB 1478 GCTCACTGATCAGCTCCCGAGGATGAGTGTCTGGCCAGCAGCTCAGCAAGGCGCTCG 1537  
QY 1740 AGAGCTGCTCTGCTGGGTGTAGGCAAGTGAAGCGGCTGCGGGCAGAGCTGAGGAGATAC 1799  
DB 1538 GGAGCTACTCTCTCTGGCGTGGGCAAGTGAAGTCTCTGAGGCGGAGCTGCTGAAGATCAC 1597

QY 1800 AGCAGCCCAAGCATGTGATGGTCTACTCTGGATCTCTCAGGATCTGTTCAACCAATCC 1859  
DB 1598 CGGTAGCCCGAAGCATGTGATGGTCCACACAGACCCCTCAGGACCTGT--CAGCCAAATCC 1655  
QY 1860 TGAGCTGACAGGGAAGCTGTGAGCGCGGAGCGGCGGAGGTGCCGACACAAAGCCCTGGA 1919  
DB 1656 AGAGCTGACAGAGGCTGTATGAGCAGCCAGCCAGCGGCTGCCAGGACACAGTCACTGA 1715  
QY 1920 CCTCGTCTTTCATGTTGGACACCTCTGCTCAGTAGGGCCCGGAGAAATTTTGTCTCAGATGA 1979  
DB 1716 CTTGCTTCTCTG-TGGATGCTCTG-CTCTGTGGACCTGAGAACTTTTGCCCAATGCA 1773  
QY 1980 GAGCTTTGTGAGAGCTGTGCTTCCCTCCAGTTTGAAGTGAACCTGAGCTGACAGAGTCGG 2039  
DB 1774 GAGCTTTCATCAGAAATGACCCCTCGGTTTGTATGTAATCTCTGATGTGACACAAGTTGG 1833  
QY 2040 CCGT 2043  
DB 1834 CCGT 1837

## RESULT 4

AAC99736

ID AAC99736 standard; cDNA; 1837 BP.

XX AAC99736;

XX 08-MAR-2001 (first entry)

XX Skin cell cDNA, SEQ ID NO: 350.

XX Mouse; skin cell; cytostatic; antiinflammatory; anti-HIV;

KW neurotropic; neuroprotective; vulnery; immunomodulatory; vaccine;

KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;

XX inflammation; neurological disease; ss.

OS Mus sp.

XX WO200069884-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000WO-NZ00075.

XX 14-MAY-1999; 9905-0312283.

XX (GENE-) GENESIS RES &amp; DEV CORP LTD.

XX Watson JD, Strachan L, Onrust R, Sleeman M, Kumble KD, Murison JG;

XX WPI; 2001-007495/01.

XX New isolated polynucleotide used in the identification of genetic disorders and encoding polypeptides used for treating inflammatory disease, cancer and neurological diseases -

PS Claim 1; Page 272-273; 352pp; English.

XX The present polynucleotide encodes a polypeptide which is expressed in mammalian skin cells. The polypeptide is useful for stimulating keratinocyte growth and motility, inhibiting the growth of cancer cells, modulating angiogenesis, inhibiting angiogenesis and vascularisation of tumours, modulating skin inflammation, stimulating the growth of epithelial cells, inhibiting the binding of human immunodeficiency virus (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and neurological diseases. The polynucleotide can be used as a marker, in the identification of genetic disorders, and for the design of oligonucleotides for examining expression patterns.

SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;

Query Match 33.1%; Score 1117.2; DB 22; Length 1837;

Best Local Similarity 79.8%; Pred. No. 3.9e-267;

Matches 1392; Conservative 0; Mismatches 343; Indels 9; Gaps 6;

```
QY 302 CGCGCTCTCTCCGCTATATCAACATGCCCCCTTTCTGTGTGCTGGAGCGGTCTGTG 361
Db 101 CGCGGCTCGCACAGTCTTAACAACATGCTCCACTTCTGCTTACCAGCCATCTACA 160
QY 362 TTTTCTCTTTTCCAGAGTCCCGCATCTCTCCCTCTCCAGGAGTCCATGTAAAGCAAG 421
Db 161 TGTCTCTGTCTTTCAGAGTGTCCCGACCATCTCTCTTCCAGGAAGTGCATGTGAACCGGG 220
QY 422 AAACCATCGGGAAGATTTCAGCTGCCAGCAAAATGATGTGCTCGCTGCAGTGGACA 481
Db 221 AGACATGGGGAGATCGCTGTGCGCCAGCAAAATTAATGTGTGCTCAGCCGCGGTGACA 280
QY 482 TCATGTTTCTGTAGATGGCTTAACAGCGTCCGGAAGGAGCTTTGAAGGTCCAAGC 541
Db 281 TCCTGTTTCTGTAGATGGCTCTCACAGCATCGGAAGGGAGCTTCGAGAGGTCCAAGC 340
QY 542 ACTTTGCCATCAGATCTGTAGCGTCTGACATCTGACATACAGCCCCGAGAGGTGAGAG 601
Db 341 GCTTTCGCATCGTGTGCTGTGATGCGCTGGACATCAGCCCTGGCAGGGTCCAGATCGGAG 400
QY 602 CATTTCAGTTTCACTTCCACCTCATCTGGAATTCCTTGGATTCATTTTCAACCCAAC 661
Db 401 CCTTGCAGTTTGGTTCACCTTCTCATCTGGAATTCCTTGGACTCTCTCAACTCGAC 460
QY 662 AGGAAGTGAAGGCAAGAAATCAAGAGGATGTTTTTCAAAGGAGGCGCACGAGACGGAAC 721
Db 461 AGGAAGTGAAGGAAAGCATCAAGGGATAGTTTTTCAAAGGTGGCGCACGAGACGGGCC 520
QY 722 TTGCTCTGAATACCTTCTCACAGAGGTTGCTTGAGGCGAGAAATGCTTCTGTGCCCC 781
Db 521 TAGCCCTGAAA---CGCCTCAGCAGAGGGTTCCCGCGAGGAGAGAAATGGCTCTGTGCCCC 577
QY 782 AGATCTCATCTCATCTGCTGGAAGTCCAGGGGAGTGGCACTCCCATCCAAGC 841
Db 578 AGATCTTATCTGTCAGCGATGGCAGTCCAGGGGCCGTGGCTCTCCCGGCTAAGC 637
QY 842 AGCTGAAGGAAGGGGTGTCACCTGTGTGTGCTGTGGGGTTCAGGTTTCCAGGTGGGAGG 901
Db 638 AGCTGAGAGAAAGGGGATCGTCTGTGTTGCCGTAGGAGTCCGTTTTTCCAGGTGGGAGC 697
QY 902 AGCTGATGACATGCGCCAGGAGCTGAGGGGAGCAGCTGCTGTGTGGCTGAGCAGGTGG 961
Db 698 AGCTGTCTCAGCTGTGGCAGTGAGCCGAAGACCGGATGCTGTGTGGCTGAGCAAGTGG 757
QY 962 AGGATGCCAACAGCGCTCTTACGACCTCAGCAGCTCGGCACTGCTCCAGCGCCA 1021
Db 758 AGGATGCCAACATGGCTCTTACGACCTCAGCAGCTCAGCAGCTCGCAGCTGACCACTGCTG 817
QY 1022 CGCCAGACTGCAGGTCGAGGCTACCCCTGTGAGCACAGGAGCTGGAGATGGTCCGGG 1081
Db 818 ATCCAGACTGCAGGTTGGAACCTCATCCCTGTGAGCGGAGGACGCTGGAGACGCTCAGGG 877
QY 1082 AGTTCGCTGCAATGCCCATGCTGGAGAGGATCGCGGGGACCTTGGCGGTGCTGGCTG 1141
Db 878 AGCTCGCTGCAATGGCTTGTGTGGAGAGGATCAAGGCAAGCAGACACTGTGTGTGCTC 937
QY 1142 CACACTCTCCCTTCTACAGCTGGAAGAGTGTTCCTTAACCCACCTGCCACCTGTGTACA 1201
Db 938 TGCCCTGTCCCTTCTACAGCTGGAGAGAGTGTTCAGACAGACACCTGCCAAGTGTACA 997
QY 1202 GGACCACTGCCAGGCCCCGTGTGACTCGCAGGCCCTGCCAAGATGGAGGACATGTGTTTC 1261
Db 998 GAACCATCTCTCCAGGCCCCGTGTGACTCCAGCCCTGCCAAGATGGAGGACCGTGCATTC 1057
QY 1262 CAGAGGACTGGAGGCTACCATGCTCTGCCCCGTGGCTTGGAGGGAGGCTTAAGT 1321
Db 1058 CAGAAGGTGTGGATAGGTACCATGCTCTGTCCCACTGGCAATTCGGAGGGAGAGTCAACT 1117
QY 1322 GTCCCTTGAAGCTGAGCTGGAATGAGGGTTCGACCTTCCTTCTTCTGCTGACAGCTCTG 1381
Db 1118 GTCCCCGAAGCTGAGCTGGAATGCAGAAATCGATGTCTCTTCTTCTGCTGACAGCTCTG 1177
```

```
QY 1382 CGGGCACCACTCTGACGGGTTCTCTCGGGGCCAAAAGTCTTCTGTGAAGCGGTTTGTGGGG 1441
Db 1178 CAGGCACCACTATTGGGGGGTTCCTCGGAGGCCAAGGCTTGTGTCAGGCGTTTGTGTCAGG 1237
QY 1442 CCGTCTGTGACGAGGACTCTCGGGCCGAGTGGGTGTGGCCACATACAGCAGGAGCTGC 1501
Db 1238 CCGTCTGTGAGGAGGACTTCCGAGCCCGCTTGGGATAGCCAGTTATGGCAGGAATCTAA 1297
QY 1502 TGTGTGCGGTG-CCTGTGGGGAGTACCAGGA-TGTGCTCTGACCTGGTCTGAGGCTCTGA 1559
Db 1298 TGTGTGCGGTGCTGTGCGGAGTACCAGCATTTGTCGCGGACCTGATCAGGAGCTTGA 1357
QY 1560 TGCATTTCCTTCCGTGTGTGGCCCCCTGACGGGAGTGCCTTGGCGGAGCGGAGCAGA 1619
Db 1358 CAGCATTTCCCTTTCAGCGGTGGCCCCGACCTTAACCGGGAGTGCCTTGTCTCCAGGTGCGAGA 1417
QY 1620 GCCTGCTTCGGAGCGCCACACAGGACAGGACGAGGACCGGCGCAGCTAGAGTGTGGTTT 1679
Db 1418 GCACGGCTTGGGAGTGCACAGAGGACTGTGTCAAGGACAGGCGCAGAGTAGTAGTTCT 1477
QY 1680 GCTCACTGAGTCACTCTCCGAGGATGAGGTTCGGGGCCACAGCGCTCACCAAGGCGCG 1739
Db 1478 GCTCACTGAATCAGCTCCAGGATGAGTGTCTGGGCCAGAGCTCACCAAGGCGCTCG 1537
QY 1740 AGAGCTGCTCTCTGCTGGGTGTAGGCGGTGAGGCGGTGCGGGCAGAGCTGAGGAGATCAC 1799
Db 1538 GGAGCTACTCTCTCGGGCGTGGCGAGTGAGATCTCTCAGGCGGAGCTGCTGAAGATCAC 1597
QY 1800 AGCAGCCCCAAACATGTGATGTCTACTCGGATCCTCAGGATCTGTTCAACCAATCCC 1859
Db 1598 CGGTAGCCCCGAGCATGTGATGTCTCACAGACACAGACCTCTCAGGACCTGT--CAGGCCAATCC 1655
QY 1860 TGAGCTGCAGGGGAAGCTGTGCAGCGCGGACAGCGGCTGGCGGACACAAAGCCCTGGA 1919
Db 1656 AGAGCTGCAGAGGAGCTATGACAGCCAGCCACCGCAGGCTGCCAGGACAGTCACTGGA 1715
QY 1920 CTTGCTTTCATGTTGGACACCTCTGCTCAGTAGGGCCCGAGAAATTTTCTCAGATGCA 1979
Db 1716 CTTGGTCTTCTCTG-TGATGCCCTCTG-CTCTGTGGGACGTGAGAAGCTTTGCCCAAAATGCA 1773
QY 1980 GAGCTTTGTGAGAGCTGTGCCCTCCAGTTTGAAGTTGAACCTTGAGCTGACAGAGTGG 2039
Db 1774 GAGCTTCATCAGGAATATGACACCTCCGCTTGTGATGTAATCTCTGATGTGACAAAGTTGG 1833
QY 2040 CTTG 2043
Db 1834 CTTG 1837
```

RESULT 5  
ABL34888

ID ABL34888 standard; cdna; 1837 BP.

XX ABL34888;

XX AC

XX 04-APR-2002 (first entry)

XX Murine cDNA isolated from skin cells SEQ ID NO: 350.

XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
developmental defect; inflammatory disease; dermatological; vulnary;  
immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;  
ss.

XX Mus sp.

XX WO200190357-A1.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-NZ00099.

XX

PR 24-MAY-2000; 2000US-206650P.  
PR 25-JUL-2000; 2000US-221232P.  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;  
XX WPI; 2002-122020/16.  
DR New polynucleotides and polypeptides encoded by the polynucleotides  
XX isolated from skin cells, useful for treating skin wounds, cancers,  
XX growth and developmental defects, inflammatory diseases, or for  
XX modulating immune responses  
XX  
XX Claim 1; Page 227-228; 466pp; English.  
XX The present invention provides the protein and coding sequences of cDNAs  
XX isolated from human, murine and rat skin cell libraries. The sequences  
XX can be used in the development of therapeutic agents useful in the  
XX treatment of skin diseases, including skin wounds, cancer, growth  
XX defects, developmental defects and inflammatory diseases. The proteins  
XX have important roles in the induction of hair growth, cell proliferation  
XX and cell-cell interaction, in maintaining tissue integrity, in wound  
XX healing and in modulating immune responses. The present sequence is a  
XX cDNA of the invention.  
XX  
XX Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;  
XX  
XX Query Match 33.1%; Score 1117.2; DB 24; Length 1837;  
XX Best Local Similarity 79.8%; Pred. No. 3.9e-267;  
XX Matches 1392; Conservative 0; Mismatches 343; Indels 9; Gaps 6;  
XX  
QY 302 CGCGGCTCTCCGTTATATCAAGATGCCCCCTTCTCGTGTGCGAGGCGGTCTG 361  
DB 101 CGCGGGGTGCACACAGGTCTAACAACATGCTCCACTTCTTACCAGCCATCTACA 160  
QY 362 TTTTCTCTGTTTCCAGATGCCCCCTCTCTCCCTCCAGGAAGTCCATGTAAACAAG 421  
DB 161 TGTCTGTTTCTTTCAGAGTGTCCCGACCATCTCTCTCAGGAAGTGCATGTAAACGGG 220  
QY 422 AAACCATCGGGAAGATTTTCAGCTGCCAGCAAAATGATGTGCTCGGCTGCAGTGGACA 481  
DB 221 AGACCATGGGAAGATCGCTGTGCCAGCAAAATTAATGTGCTCAGCGCGGTGCACA 280  
QY 482 TCATGTTCTGTAGATGGGTCTAAGAGGTGCGGAAAGAGGAGCTTTGAAAGTCCAAGC 541  
DB 281 TCCTGTGTTCTGTAGATGGTCTCACAGCATCGGGAAGGAGCTTCGAGAGGTCCAAGC 340  
QY 542 ACTTTGCCATCAGATCTGTGACGGTCTGACATCAGCCCGAGAGGTGTCAGATGGGAG 601  
DB 341 GCTTCGCCATCGCTGCTGTGATGCCCTGGACATCAGCCCTGGCAGGTGTCAGATCGGAG 400  
QY 602 CATTCCAGTTTCAGTCTCCACTCTCATCTGGAATTCCTCCCTTGGATTCATTTTCAACCCAAC 661  
DB 401 CCTTGCAGTTTGTCTCACTCTCATCTGGAATTCCTCCCTTGGACTCCTTCTCAACTCGAC 460  
QY 662 AGGAAGTGAAGCAAGATCAAGAGGATGTTTTCAAAGAGGCGGACGAGAGCGGAAC 721  
DB 461 AGGAAGTGAAGCAAGATCAAGAGGATGTTTTCAAAGGTTGGCGCACCGAGAGCGGCC 520  
QY 722 TTCTCTGTAATACCTTCTGCACAGAGGTGCTCGAGGACAGAAATGCTCTGTGCCCC 781  
DB 521 TAGCCCTGAAA---CGCTGAGCAGAGGTGTTCCCGGAGGAGCAATGCTCTGTGCCCC 577  
QY 782 AGATCTCATCATCTGCTAGTGGGAAGTCCCGAGGGGATGTGGCACTGCCATCCAAGC 841  
DB 578 AGATTTCTATCATCTGACAGGATGGCAAGTCCCGAGGGCCCGTGGCTCTCCCGGCTAAGC 637  
QY 842 AGCTGAAGAAAGGGGTGCTACTGTGTGCTGTGGGGGTGTCAGGTTTCCAGGTGGGAGG 901  
DB 638 AGCTGAAGAAAGGGGATCGTCTGCTGTTGCCGTAGGAGTCCGTTTTCAGGTGGGAGC 697  
QY 902 AGCTGATGACCTGGCCAGGAGCCTAGAGGGGACAGCATGCTGTGTTGGTGAAGAGTGG 961

DB 698 AGCTGCTACGCTGGCGAGTGGAGCGGACCGGATGTGCTGTTGGCTGAGCAAGTGG 757  
QY 962 AGGATGCCACCAAGCGCTCTTTCAGCACCCCTCAGCAGCTCGGCCATCTGCTCCAGCGCCA 1021  
DB 758 AGGATGCCACCAATGGCTCTCCTCAGCACCCCTCAGCAGCTCCGACTCTGCACCACTGCTG 817  
QY 1022 CGCCAGACTGACAGGTGTCAGGCTCACCCTGTGAGCACAGGACGTCGAGATGGTCCGGG 1081  
DB 818 ATCCAGACTGACAGGTGGAACCTCATCCCTGTGAGCGGAGGACGCTGGAGAGCCGTGAGG 877  
QY 1082 AGTTCGCTGGCAATGCCCATGCTGAGAGGATCGCGGGGACCCCTGCGGTGCTGGCTG 1141  
DB 878 AGCTGCTGGCAATGCCCTGTGCTGAGAGGATCAAGGCAAGCAGACACTGTGCTGGCTC 937  
QY 1142 CACACTGTCCCTTCTACAGCTGGAAGAGAGTGTTCCTAACCCACCCTGCCACCTGCTACA 1201  
DB 938 TGCCCTGTCCCTTCTACAGCTGGAAGAGAGTGTTCAGACACACACCCCTGCCAATGCTACA 997  
QY 1202 GGACCACTGCCAGGCCCTGTGACTCGAGCCCTGCCAGATGAGGACACATGTGTTTC 1261  
DB 998 GAACCATCTGTCCAGGCCCTGTGACTCCAGCCCTGCCAATGAGGACACGTCGATTC 1057  
QY 1262 CAGAAGGACTGGAGGCTTACCAGTGTCTGCGCGCTGCGCTTTTGAGGGGAGGCTAACT 1321  
DB 1058 CAGAAGGTGGATAGTAGTACACTGTCTGCCCACCTGCGATTCGGAGGGAGTCAACT 1117  
QY 1322 GTGCCCTGAAGCTGAGCTGGAATGCAAGGTGCGACCTCTCTTCTGCTGCGACAGCTCTG 1381  
DB 1118 GTGCCCCGAAGCTGAGCCCTGGAATGCAGAATCGATGTCTCTCTCTGCTGGACAGTCTG 1177  
QY 1382 CGGGACCACTCTGGAGGCTTCTCGCGGCCCAAGTCTCTGTAAGCGGTTTGDCGGG 1441  
DB 1178 CAGGACCACTATTTGGGGGCTTCCGGAGGGCCAAAGCCCTTGTCAAGCCCTTGTGCAGG 1237  
QY 1442 CGTGTGTGAGGAGGACTCTCGGGCCCGAGTGGTGTGCGCACATACAGCAGGAGCTGC 1501  
DB 1238 CCGTGTGAGGAGGACTCCCGAGCCCGGTGGGATAGCCAGTTATGCGAGGAATCTAA 1297  
QY 1502 TGTGTGCGGTG-CCTGTGGGGAGTACCAGGA-TGTGCTCTGACCTGCTGTGAGCCCTCGA 1559  
DB 1298 TGTGTGCGGTGCTCTCGGGGAGTACCAGCATGTGTGCGGACCTCATCAGGAGCCCTGA 1357  
QY 1560 TGGCATTTCCCTTCGCTGGTGGCCCACTGAGCGGAGTGCCTTGGCGGACGCGGACAGA 1619  
DB 1358 CAGCATTTCCCTTCAGCGGTGGCCCACTTAACCGGAGTGCCTTCTCCAGGTGCGAGA 1417  
QY 1620 GCGTGTGCTTGGGAGCGCCACAGGACAGGCCAGGACCGGCCACGTAGACTGTGCTTTT 1679  
DB 1418 GCAGCGCTTGGGAGTGCAGCAGGACTGTCAGGACAGCCACGACGAGTAGTACTTCT 1477  
QY 1680 GCTCACTGATCACACTCCGAGGATGAGTGTGGGGCCCGAGCGCTCAGCAAGGCGGG 1739  
DB 1478 GCTCACTGAATCAGCTCCAGGATGAGTGTGCGGCCAGCAGCTCAGCAAGGCGTGC 1537  
QY 1740 AGAGCTGCTCTGCTGGTGTAGGAGTGTGAGCGGCTGCGGGCAGAGCTGGAGAGATCAC 1799  
DB 1538 GGAGCTACTCTCTCGGCTGGGAGTGTGAGATCTTGTGAGGCGGAGTGTGAGAGTCAAC 1597  
QY 1800 AGGACGCCCCAAAGCATGTGATGTGTCTACTCGGATCTCTCAGGATCTGTCAACCAATCCC 1859  
DB 1598 CGGTAGCCCCGAAGCATGTGATGTGTCCACAGACACCTCTCAGGACCTGT--CAGCCAATCC 1655  
QY 1860 TGAGCTGACGGGAAAGCTGTGAGCGGACGCGGCGAGGCTGCGGACACAAAGCCCTGGA 1919  
DB 1656 AGAGCTGACAGAGGAGCTATGACCGCAGCCACGCGCCAGGCTGCGCAGGACAGTCACTGGA 1715  
QY 1920 CCTCGTCTTCACTGTGGACACCTCTGCTCAGTAGGGCCGAGAAATTTTCTCAGATGCA 1979  
DB 1716 CTTGGTCTTCTCTG-TGAGTGCCTCTG-CTCTGTGGGACGTTGAGAACTTTGCCCAATGCA 1773  
QY 1980 GAGCTTTGTGAGAAGCTGTGCCCTCCAGTTTGTAGGTGAACTTGTGAGCTGACAGAGTCCG 2039

Db 1774 GAGCTTCATCAGGAATGCACCCCTCCGTTTGATGTGAATCCTGATGTCGACACAAGTTGG 1833  
QY 2040 CCTG 2043  
|||||  
Db 1834 CCTG 1837  
RESULT 6  
AAS81529/c  
ID AAS81529 standard; cDNA; 406 BP.  
XX AC  
AC AAS81529;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
XX  
DE DNA encoding novel human diagnostic protein #17333.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG17342.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX  
PS Claim 1; SEQ ID NO 17333; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 406 BP; 80 A; 112 C; 107 G; 107 T; 0 other;  
S0

Query Match 10.1%; Score 340; DB 23; Length 406;  
Best Local Similarity 90.7%; Pred. No. 2.1e-74;  
Matches 390; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 2482 AGCCCGTGCATGAATGAGGCGAGCTCGTCTCTGCAGAAATGGAGCTACCGCTGCAAGTGT 2541  
|||||  
Db 406 AGCCCGTGCATGAATGAGGCGAGCTCGTCTCTGCAGAAATGGAGCTACCGCTGCAAGTGT 347  
|||||  
QY 2542 CGGGATGGCTGGGAGGCGCGCCCACTGCGAGAACCGTGAAGTGGAGCTCTTGCCTCTGTATGT 2601  
|||||  
...Db 346 CGGGATGGCTGGGAGGCGCGCCCACTGCGAGAACCGTGAAGTGGAGCTCTTGCCTCTGTATGT 313  
|||||  
QY 2602 GTGAGCCAGGATGGATTCTTTGAGACGCCCTGTAGGCACATGGCTCCCGTGCAGGAGGGC 2661  
|||||  
Db 312 -----GATTCTTGAGACGCCCTGTAGGCACATGGCTCCCGTGCAGGAGGGC 267  
|||||  
QY 2662 AGCAGCGGTACCCCTCCAGCAACTACAGAGAAGCGCTGGGCACTGAAATGGTCCCTACC 2721  
|||||  
Db 266 AGCAGCGGTACCCCTCCAGCAACTACAGAGAAGCGCTGGGCACTGAAATGGTCCCTACC 207  
|||||  
QY 2722 TTCTGGAAATGCTGTGCTGCCCGAGGTCCTTAGAATCTCTGCTTCCCGCGTGGCCAGGACCA 2781  
|||||  
Db 206 TTCTGGAAATGCTGTGCTGCCCGAGGTCCTTAGAATCTCTGCTTCCCGCGTGGCCAGGACCA 147  
|||||  
QY 2782 CTATTCTCACTGAGGAGGAGGATGTCCCAACTGCAGCCATGCTGCTTAGAGACAAGAAA 2841  
|||||  
Db 146 CTATTCTCACTGAGGAGGAGGATGTCCCAACTGCAGCCATGCTGCTTAGAGACAAGAAA 87  
|||||  
QY 2842 GCAGCTGATGTCACCCACAAACGATGTTGTTGAAAGTTTTTGATGTGTAAGTAATACCC 2901  
|||||  
Db 86 GCAGCTGATGTCACCCACAAACGATGTTGTTGAAAGTTTTTGATGTGTAAGTAATACCC 27  
|||||  
QY 2902 ACTTTCGTGA 2911  
|||||  
Db 26 ACTTTCGTGA 17  
|||||  
RESULT 7  
ABQ16108/c  
ID ABQ16108 standard; DNA; 651 BP.  
XX  
XX ABQ16108;  
AC  
XX  
DT 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 2699.  
DE  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200218632-A2.  
PN  
XX 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001WO-EP10074.  
PF  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
XX 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPIG-) EPICENOMICS AG.  
PA  
XX  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI WPI; 2002-371829/40.  
XX  
XX  
DR  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
PS  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a







PN W0200177389-A2.

XX 18-OCT-2001.

XX 04-APR-2001; 2001WO-US11128.

XX 05-APR-2000; 2000US-195106P.

XX (INCY-) INCYTE GENOMICS INC.

XX Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;

PI Tai J;

XX WPI; 2002-010925/01.

XX Composition useful for diagnosis of conditions, disorders or diseases

PT associated with atherosclerosis, comprises several polynucleotides that

PT are differentially expressed in foam cell development

XX Claim 1; Page 84-85; 315pp; English.

XX The present invention relates to the isolation of human polynucleotide

CC sequences that are differentially expressed during foam cell

CC differentiation. The polynucleotide sequences of the invention or a

CC composition comprising these polynucleotides are useful as a high

CC throughput method for detecting altered expression of one or more

CC polynucleotides in a sample. The polynucleotides can be used in the

CC diagnosis of disorders associated with foam cell development such as

CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as

CC coronary artery disease. The polynucleotide sequences can also be used

CC as PCR primers and probes. The polynucleotides of the invention are also

CC useful in gene therapy. AAS94746-AA595021 represent the human

CC polynucleotide sequences of the invention which are differentially

CC expressed during foam cell differentiation.

XX Sequence 4151 BP; 889 A; 1089 C; 1163 G; 928 T; 82 other;

SQ

Query Match 2.9%; Score 97.8; DB 24; Length 4151;

Best Local Similarity 50.6%; Pred. No. 7e-14;

Matches 265; Conservative 0; Mismatches 253; Indels 6; Gaps 1;

QY 422 AACCATCGGAGATTTCAGCTCCAGCAAAATGATGTGCTCGCTCCAGTGCA 481

DB 1128 ACAGCGACGGACCACTGCAATGCTGACAGTGTGCTGAGCGCTCGGCCACTGACC 1187

QY 482 TCATGTTTCTGTAGATGGTCTAACAGCGTCCGGAAGGAGGCTTTGAAAGGTCCAAGC 541

DB 1188 TGGTCTTCTTACGGATCCAGAGTGTGAGCCAGAGAACCTTGACCTGGTGAAGA 1247

QY 542 ACTTTGCCATCAGCTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCCAGAGTGGGAG 601

DB 1248 AGTTTCATCAGTCAGATCGTGATACGCTGGAGGTGTGACAGCAAGCTGGCCAGGTGGGC 1307

QY 602 CATTCAGTTTCAGTTCACCTCCTCATCTGGAATTCCTTGGATTTCATTTCAACCCCAAC 661

DB 1308 TGGTGCAGTACTCAAGCTCTGTGCGCCAGGAGTTCCCTCCCTTCCACACCAAGA 1367

QY 662 AGGAAGTGAAGGCAAGAAATCAAGAGGATGTTTTCAGAGGAGGCGCACGAGACGGAAC 721

DB 1368 AGGACATCAAGCGGCTGTCCGGAATATGCTTACATGGAGAGGGGCAATGACCCGGG 1427

QY 722 TTGCTCTGAATACCTTCTCCACAGAGGGTTCCTTCAGAGGAGCAAAATGCT-----TCTG 775

DB 1428 CTGCTCTCAAGTACCTCATTTGACAATTCCTTCACTGTGTCCAGTGTGCTAGGCGCCGGG 1487

QY 776 TGCCCCAGATCCCATCATCTGCTACTGATGGAGTCCAGGGGGATGTGGCACTGCCAT 835

DB 1488 CCCAGAAGGGGGGATTTGCTTCACTATGCGCGGAGGAGGACTACATTAATGATGCTG 1547

QY 836 CCAAGCAGCTGAAGGAAAGGGGTGCTACTGTGTTTGTCTGTGGGGTTCAGGTTTCCCAAGT 895

DB 1548 CCAGAAGGCCAAGAGCTCGGCTTTAAGATGTTTGTGTGGGTGTGGCAATGCCCTGG 1607

QY 896 GGGAGGAGTGCATGTCACCTGCCAGCGAGCTAGAGGCGCAGCAC 939

DB 1608 AGGATGAGCTGAGGAAATAGCCTCAGAGCTGTGGCAGAGCAC 1651

III IIIII I I III IIIII I IIIII

RESULT 13

AA00664

ID AAX00664 standard; DNA; 1668 BP.

XX AAX00664;

XX 25-MAR-1999 (first entry)

XX Human secreted protein gene 54 clone HE8EM69.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX W09842738-A1.

XX 01-OCT-1998.

XX 19-MAR-1998; 98WO-US05311.

XX 30-MAY-1997; 97US-0050937.

XX 21-MAR-1997; 97US-0041276.

XX 21-MAR-1997; 97US-0041277.

XX 21-MAR-1997; 97US-0041281.

XX 21-MAR-1997; 97US-0042344.

XX 30-MAY-1997; 97US-0048069.

XX 30-MAY-1997; 97US-0048094.

XX 30-MAY-1997; 97US-0048095.

XX 30-MAY-1997; 97US-0048096.

XX 30-MAY-1997; 97US-0048099.

XX 30-MAY-1997; 97US-0048131.

XX 30-MAY-1997; 97US-0048135.

XX 30-MAY-1997; 97US-0048154.

XX 30-MAY-1997; 97US-0048160.

XX 30-MAY-1997; 97US-0048186.

XX 30-MAY-1997; 97US-0048187.

XX 30-MAY-1997; 97US-0048188.

XX 30-MAY-1997; 97US-0048350.

XX 30-MAY-1997; 97US-0048351.

XX 30-MAY-1997; 97US-0048352.

XX 30-MAY-1997; 97US-0048355.

XX 05-AUG-1997; 97US-0054804.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;

XX Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;

XX Rosen CA, Ruben SM, Shi Y, Young P;

XX WPI; 1999-070066/06.

XX P-PSDB; AAW67860.

XX New isolated human genes and the secreted polypeptides they encode

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX Claim 1; Page 215-216; 385pp; English.

XX This sequence represents a nucleic acid molecule which encodes a

CC secreted human protein. The gene number, and the clone it is derived

CC from, are detailed in the descriptor line. The gene can be used to



Db 414 AGCGCAGGTCACGCCCGCTTAAACGTGCTCCTGTTGGCTGCTCTTCACGG 473  
QY 2249 GCGGAGAGCGGACGAGCGCTTCTCTGCCAGAGCTGAGGAACAATGGCATCT 2308  
Db 474 ATGCCCTCCAGGATGACATCTCGGTGGGAGCGCGCCAGGAGGAGGATCG 533  
QY 2309 CTGCTTTGGTGGGCGTGGGCGCTTCTCTAACTGAGGTCTGCGGAGGCTTGC 2363  
Db 534 TCATGTACGCGCTGGCGTGGCAAGCGGTGGAGGCGGAGCTGCGGAGATCGC 588

RESULT 15  
AAH17154  
ID AAH17154 standard; cDNA; 1167 BP.  
XX AC AAH17154;  
XX DE 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:16503.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 16503; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.  
XX Sequence 1167 BP; 203 A; 337 C; 424 G; 203 T; 0 other;  
SQ  
Query Match 2.2%; Score 74.2; DB 22; Length 1167;  
Best Local Similarity 49.1%; Pred. No. 3.1e-08;  
Matches 233; Conservative 0; Mismatches 233; Indels 9; Gaps 1;  
QY 1898 GGTGCCGACACAAAGCCCTGGACCTCTTCTTATGTTGGACACCTCTGCCTCAGTAGGC 1957  
Db 114 GGTGCCGGAAGCCACGCTGGACCTTCTTCTGCTGTTGATGGCTCCAAGAGCGTGGCTC 173  
QY 1958 CCGAGAAATTTTGTCTAGATGACAGCTTTGTGAGAAGCTGTGCCCTCCAGTTTGGAGTGA 2017  
Db 174 CACAAAATCTTCGAGCTAGTGAAGCGCTTCGTGAACACAGATTGTGGACTTCTCTAGATGT 233  
QY 2018 ACCTGACGTGACACAGGTGCGGCTGGTGTATGTCACGACGAGTGCAGTCTGCTTCCG 2077  
Db 234 CCCCCGAGGACGCGGCTGGTGTGAGTTCTCGAGCCGCTGCGGACCCAGTTCC 293  
QY 2078 GGCTGGACACCAAAACCCACCCGCGCTGCGATGCTGCGGGCCATTAGCCAGGCCCTTACC 2137  
Db 294 CTCTGGTCTCTAGCGCACCCAGCCAGCTGAAGCAGCGGCTCTGCGCGTGGAGTACA 353  
QY 2138 TAGGTGGGTGGGCTCAGCCGCGCACCCCTGTGTCACATCTATGACAAAGTGTATGACCG 2197  
Db 354 TGGAAACGCGCACCATGACAGGGCTGGCTTGGGACATGTTGGGACACATGTTGAGCAGCTTCTCCG 413  
QY 2198 TCCAGAGGGTGGCGGCTGGTGTCTCCCAA-----AGCTGTGGTGGTCTCACAG 2248  
Db 414 AGCGCAGGTGACGCGCGCTTAACTGTCGCTGTTGGCTGCTTCTTTCACGG 473  
QY 2249 GCGGAGAGCGGACGAGGATGACGCGCTTCTCTGCCAGAAAGCTGAGGAACAATGGCATCT 2308  
Db 474 ATGCCCTCCAGGATGACATCTCGGTGGGACGCGCGCCAGGAGGAGGAGGATCG 533  
QY 2309 CTGCTTTGGTGGGCGTGGGCGCTTCTCTAACTGAGGTCTGCGGAGGCTTGC 2363  
Db 534 TCATGTACGCGCTGGCGTGGCAAGCGGTGGAGGCGGAGCTGCGGAGATCGC 588

Search completed: May 5, 2003, 02:27:15  
Job time : 493 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 02:27:24 ; Search time 248 Seconds  
(without alignments)  
16080.529 Million cell updates/sec

Title: us-09-930-020a-1

Perfect score: 3375

Sequence: 1 gacagtggttcgcgctgcac.....tgtccaccttgaggtcttc 3375

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PTCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PTCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3375	100.0	3375	9	US-09-930-020a-1
2	1117.2	33.1	1837	9	US-10-152-661-350
3	1117.2	33.1	1837	9	US-09-866-050A-350
4	74.6	2.2	1668	9	US-09-984-245-64
5	74.6	2.2	1668	9	US-09-966-262-64
6	74.6	2.2	1668	9	US-09-983-966-64
7	74.6	2.2	1668	9	US-10-143-090-64
8	73.6	2.2	3447	9	US-10-000-512-9
9	73.6	2.2	3449	9	US-09-905-291A-33
10	73.6	2.2	3449	9	US-09-902-853-33
11	73.6	2.2	3449	9	US-09-907-824-33
12	73.6	2.2	3449	9	US-09-907-841-33
13	73.6	2.2	3449	9	US-09-904-011-33
14	73.6	2.2	3449	9	US-10-028-072-293
15	73.6	2.2	3449	9	US-09-906-742-33
16	73.6	2.2	3449	9	US-10-121-049-293
17	73.6	2.2	3449	9	US-10-123-904-293
18	73.6	2.2	3449	9	US-10-140-470-293
19	73.6	2.2	3449	9	US-09-906-838-33

20	73.6	2.2	3449	9	US-09-907-613-33	Sequence 33, Appl
21	73.6	2.2	3449	9	US-09-907-942-33	Sequence 33, Appl
22	73.6	2.2	3449	9	US-10-175-746-293	Sequence 293, App
23	73.6	2.2	3449	9	US-10-176-918-293	Sequence 293, App
24	73.6	2.2	3449	9	US-10-176-921-293	Sequence 293, App
25	73.6	2.2	3449	9	US-10-137-865-293	Sequence 293, App
26	73.6	2.2	3449	9	US-10-140-474-293	Sequence 293, App
27	73.6	2.2	3449	9	US-09-904-820-33	Sequence 33, Appl
28	73.6	2.2	3449	9	US-09-904-859-33	Sequence 33, Appl
29	73.6	2.2	3449	9	US-09-909-204-33	Sequence 33, Appl
30	73.6	2.2	3449	9	US-10-142-431-293	Sequence 293, App
31	73.6	2.2	3449	9	US-10-143-114-293	Sequence 293, App
32	73.6	2.2	3449	9	US-09-904-786-33	Sequence 33, Appl
33	73.6	2.2	3449	9	US-09-906-646-33	Sequence 33, Appl
34	73.6	2.2	3449	9	US-09-906-700-33	Sequence 293, App
35	73.6	2.2	3449	9	US-10-140-002-293	Sequence 33, Appl
36	73.6	2.2	3449	9	US-09-902-903-33	Sequence 33, Appl
37	73.6	2.2	3449	9	US-09-903-749A-33	Sequence 33, Appl
38	73.6	2.2	3449	9	US-09-903-786-33	Sequence 33, Appl
39	73.6	2.2	3449	9	US-10-142-419-293	Sequence 293, App
40	73.6	2.2	3449	9	US-09-902-736-33	Sequence 33, Appl
41	73.6	2.2	3449	9	US-09-904-119-33	Sequence 33, Appl
42	73.6	2.2	3449	9	US-09-904-956-33	Sequence 33, Appl
43	73.6	2.2	3449	9	US-09-907-794-33	Sequence 33, Appl
44	73.6	2.2	3449	9	US-10-123-262-293	Sequence 293, App
45	73.6	2.2	3449	9	US-10-142-423-293	Sequence 293, App

ALIGNMENTS

RESULT 1

US-09-930-020a-1

: Sequence 1, Application US/09930020A

: Publication No. US20030077568A1

: GENERAL INFORMATION:

: APPLICANT: Gish, Kurt C.

: APPLICANT: Mack, David H.

: APPLICANT: Wilson, Keith E.

: APPLICANT: Pos Biotechnology, Inc.

: TITLE OF INVENTION: Methods of Diagnosis of Colorectal Cancer, Compositions

: TITLE OF INVENTION: and Methods of Screening for Colorectal Cancer

: TITLE OF INVENTION: Modulators

: FILE REFERENCE: 018501-0031000S

: CURRENT APPLICATION NUMBER: US/09/930,020A

: CURRENT FILING DATE: 2001-08-14

: PRIOR APPLICATION NUMBER: US 09/663,733

: PRIOR FILING DATE: 2000-09-15

: NUMBER OF SEQ ID NOS: 3

: SOFTWARE: Patentin Ver. 2.1

: SEQ ID NO 1

: LENGTH: 3375

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: OTHER INFORMATION: CBF9

: FEATURE:

: NAME/KEY: CDS

: LOCATION: (328)..(2751)

: OTHER INFORMATION: CBF9

US-09-930-020a-1

Query Match 100.0%; Score 3375; DB 9; Length 3375;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGTGTTCGCGCTGCACCGCTCGAGCTGGTGCACCGCTAGCACTGAAGTACTT 60

Db 1 GACAGTGTTCGCGCTGCACCGCTCGAGCTGGTGCACCGCTAGCACTGAAGTACTT 60

QY 61 TTTTATTTGACAGACCTGGGCGGATGCCGCTTTAAAAACGCGAGGGGCTCTATGCACCTC 120

Db 61 TTTTATTTGACAGACCTGGGCGGATGCCGCTTTAAAAACGCGAGGGGCTCTATGCACCTC 120

Qy	121	CCTGGCGGTAGTTCCTCCGACCTCAGCCGGGTGCGGCTGCTCCGCCCTCTCCGAGGAG	180
Db	121	CCTGGCGGTAGTTCCTCCGACCTCAGCCGGGTGCGGCTGCTCCGCCCTCTCCGAGGAG	180
Qy	181	ACAAACAGGTGTCCCAAGTGGCAGCGCGCCCGGGGCCCTCCTGTGATCCCGTAGCG	240
Db	181	ACAAACAGGTGTCCCAAGTGGCAGCGCGCCCGGGGCCCTCCTGTGATCCCGTAGCG	240
Qy	241	CCCCCTGGCCCGAGCGCCCGCGGCTCTGTGAGTAGAGCCCGCGGACCCGAGCGCTGG	300
Db	241	CCCCCTGGCCCGAGCGCGCCCGGCTCTGTGAGTAGAGCCCGCGGACCCGAGCGCTGG	300
Qy	301	TCGCCGCTCTCCTTCGGTTATATCAACATGCCCCCTTTCCTGTGTGCTGGAGCGCTGT	360
Db	301	TCGCCGCTCTCCTTCGGTTATATCAACATGCCCCCTTTCCTGTGTGCTGGAGCGCTGT	360
Qy	361	GTTTTTCTGTGTTTCCAGAGTGGCCCATCTCTCCCTCTCCAGGAAGTCCATGTAAACAA	420
Db	361	GTTTTTCTGTGTTTCCAGAGTGGCCCATCTCTCCCTCTCCAGGAAGTCCATGTAAACAA	420
Qy	421	GAACCAATCGGGAAGATTTCAGCTGCCAGCAAAATGATGTGGTGCTCGGCTGCAGTGGAC	480
Db	421	GAACCAATCGGGAAGATTTCAGCTGCCAGCAAAATGATGTGGTGCTCGGCTGCAGTGGAC	480
Qy	481	ATCATGTTCTCTGTACATGGGTCTAACAGCGTCCGGAAGGAGCTTTGAAAGTCCAAG	540
Db	481	ATCATGTTCTCTGTACATGGGTCTAACAGCGTCCGGAAGGAGCTTTGAAAGTCCAAG	540
Qy	541	CACTTTCCCATCACAGTCTGTGACGCTCTGGACATCAAGCCCGAGAGGGTCAAGTGGGA	600
Db	541	CACTTTCCCATCACAGTCTGTGACGCTCTGGACATCAAGCCCGAGAGGGTCAAGTGGGA	600
Qy	601	GCATTTCAAGTTCAGTTCCACTCCTCATCTGGAAATCCCTTGGATTCATTTCAACCAA	660
Db	601	GCATTTCAAGTTCAGTTCCACTCCTCATCTGGAAATCCCTTGGATTCATTTCAACCAA	660
Qy	661	CAGGAAGTGAAGCAAGAATCAAGAGGATGTTTTTCAAGAGGGCGCAGGAGACGGAA	720
Db	661	CAGGAAGTGAAGCAAGAATCAAGAGGATGTTTTTCAAGAGGGCGCAGGAGACGGAA	720
Qy	721	CTTGTCTCTGAAATACCTTCTGCACAGAGGGTTGCCGTGAGCGCAAAATGCTTCTGTGCC	780
Db	721	CTTGTCTCTGAAATACCTTCTGCACAGAGGGTTGCCGTGAGCGCAAAATGCTTCTGTGCC	780
Qy	781	CAGATCTCTATCATCTGTCACTGATGGAAATCCCAAGGGGATGTGGCACTGCCATCCAAG	840
Db	781	CAGATCTCTATCATCTGTCACTGATGGAAATCCCAAGGGGATGTGGCACTGCCATCCAAG	840
Qy	841	CAGCTGAAGAAAGGGGTGTCACTGTGTTGCTGTGGGGGTCAAGTTTCCAGGTGGGAG	900
Db	841	CAGCTGAAGAAAGGGGTGTCACTGTGTTGCTGTGGGGGTCAAGTTTCCAGGTGGGAG	900
Qy	901	GAGCTGCATCACTGGCCACGAGCTAGAGGGCAGACGTGCTGTTGGCTGAGCAGGTG	960
Db	901	GAGCTGCATCACTGGCCACGAGCTAGAGGGCAGACGTGCTGTTGGCTGAGCAGGTG	960
Qy	961	GAGGATGCCACCAAGCCCTCTTTCAGCACCTCTCAGCAGCTCGGCCATCTGCTCCAGCGCC	1020
Db	961	GAGGATGCCACCAAGCCCTCTTTCAGCACCTCTCAGCAGCTCGGCCATCTGCTCCAGCGCC	1020
Qy	1021	ACGCCAGACTGCAGGTCCAGGCTCACCCCTGTGAGCACAGGACGCTGGAGATGGTCCGG	1080
Db	1021	ACGCCAGACTGCAGGTCCAGGCTCACCCCTGTGAGCACAGGACGCTGGAGATGGTCCGG	1080
Qy	1081	GAGTTCTGCTGGCAATGCCCATGCTGGAGAGGATCGCGCGGACCCCTTGCGGTGTGGCT	1140
Db	1081	GAGTTCTGCTGGCAATGCCCATGCTGGAGAGGATCGCGCGGACCCCTTGCGGTGTGGCT	1140
Qy	1141	GCACATGTCCCTTCTACAGCTGGAAGAGATGTTCTTAACCCACCCCTGCCACCTGCTAC	1200
Db	1141	GCACATGTCCCTTCTACAGCTGGAAGAGATGTTCTTAACCCACCCCTGCCACCTGCTAC	1200

Qy	1201	AGNACACCTGCCCAGGCCCTTGTA	CTCGACGAGCCCTGCCAGATGAGGCACATGTGTT	1261
Db	1201	AGGACCACCTGCCACGCCCCCTTG	ACTGACCTGCGAGCCCTGCCAGATGAGGCACATGTGTT	1260
Qy	1261	CCAGAAGGACTGCACGGCTACCA	GTGCTCTGCCCGCTGGCCTTTGGAGGGAGGCGCTAAC	1320
Db	1261	CCAGAAGGACTGCACGGCTACCA	GTGCTCTGCCCGCTGGCCTTTGGAGGGAGGCGCTAAC	1320
Qy	1321	TGTGCCCTGAAGCTGAGCCTGGA	ATGCAAGGTGCGACCTCTCTTCCTGCTGGACAGCTCT	1380
Db	1321	TGTGCCCTGAAGCTGAGCCTGGA	ATGCAAGGTGCGACCTCTCTTCCTGCTGGACAGCTCT	1380
Qy	1381	CGGGGCACCACTCTGACGGCTCT	TGCGGGCCCAAGTCTTCGTGAAGCGGTGTGTCGG	1440
Db	1381	CGGGGCACCACTCTGACGGCTCT	TGCGGGCCCAAGTCTTCGTGAAGCGGTGTGTCGG	1440
Qy	1441	GCCTGTGTAGCGAGGACTCTGG	GGCCCGAGTGGTGTGCCACATACACGAGGAGCTG	1500
Db	1441	GCCTGTGTAGCGAGGACTCTGG	GGCCCGAGTGGTGTGCCACATACACGAGGAGCTG	1500
Qy	1501	CTGTGTGCGGTGCTGTGGGGAG	TACCAAGATGTGCTGACCTGTGTGGAGCCTCGAT	1560
Db	1501	CTGTGTGCGGTGCTGTGGGGAG	TACCAAGATGTGCTGACCTGTGTGGAGCCTCGAT	1560
Qy	1561	GGCATTCCTTCGTGTGGCCCA	CCCTGTACGGGCAGTGCCTGCGCAGCGGCGACAG	1620
Db	1561	GGCATTCCTTCGTGTGGCCCA	CCCTGTACGGGCAGTGCCTGCGCAGCGGCGACAG	1620
Qy	1621	CGTGGCTTCGGGAGCGCCACCA	GGACAGCCAGGACCGGCCACGTAGAGTGGTGGTTTG	1680
Db	1621	CGTGGCTTCGGGAGCGCCACCA	GGACAGCCAGGACCGGCCACGTAGAGTGGTGGTTTG	1680
Qy	1681	CTCACTGAGTCACACTCCGAG	NATGAGTTGCGGGCCAGCGCTCAGCAAGGCGCGA	1740
Db	1681	CTCACTGAGTCACACTCCGAG	NATGAGTTGCGGGCCAGCGCTCAGCAAGGCGCGA	1740
Qy	1741	GAGCTGCTCTGCTGGGTGTAG	CGACAGTGAAGCCGTGCGGGCAGAGCTGAGAGGATCACA	1800
Db	1741	GAGCTGCTCTGCTGGGTGTAG	CGACAGTGAAGCCGTGCGGGCAGAGCTGAGAGGATCACA	1800
Qy	1801	GGCAGCCCAAGCATGTGATGCT	TACTCGGATCTCAGGATCTGTTCAACCAATCCCT	1860
Db	1801	GGCAGCCCAAGCATGTGATGCT	TACTCGGATCTCAGGATCTGTTCAACCAATCCCT	1860
Qy	1861	GAGCTGACGGGGAAGCTGTGCA	CGCCGACAGCGCCAGGGTGCCGACACAAGCCCTGGAC	1920
Db	1861	GAGCTGACGGGGAAGCTGTGCA	CGCCGACAGCGCCAGGGTGCCGACACAAGCCCTGGAC	1920
Qy	1921	CTCGTCTTCATGTTGACACCT	CTGCCTCAGTAGGGCCCGAGAAATTTGCTCAGATCGAG	1980
Db	1921	CTCGTCTTCATGTTGACACCT	CTGCCTCAGTAGGGCCCGAGAAATTTGCTCAGATCGAG	1980
Qy	1981	AGCTTTTGAGAAAGCTGTGCC	TCCAGTTGAGGTGAACCTGACGTGACACAGGTGCGC	2040
Db	1981	AGCTTTTGAGAAAGCTGTGCC	TCCAGTTGAGGTGAACCTGACGTGACACAGGTGCGC	2040
Qy	2041	CTGTGTGTGATGGACCGAGGT	GCAGATGCTCTTCGGGTGGACACAAACCCACCCGG	2100
Db	2041	CTGTGTGTGATGGACCGAGGT	GCAGATGCTCTTCGGGTGGACACAAACCCACCCGG	2100
Qy	2101	GCTCGATGCTCGGGCCATTAG	CCAGGCCCCCTACCTAGGTGGGTGGCTCAGCCGCG	2160
Db	2101	GCTCGATGCTCGGGCCATTAG	CCAGGCCCCCTACCTAGGTGGGTGGCTCAGCCGCG	2160
Qy	2161	ACCGCCCTGCTGCACATCTAT	GACAAAGTGATGACCGTCCAGAGGGGTGCCCGCCCTGGT	2220
Db	2161	ACCGCCCTGCTGCACATCTAT	GACAAAGTGATGACCGTCCAGAGGGGTGCCCGCCCTGGT	2220
Qy	2221	GTCCCCAAAGCTGTGTGTGCT	CACAGCGGGAGCGGCAGAGGATGACGCCGTTCCCT	2280
Db	2221	GTCCCCAAAGCTGTGTGTGCT	CACAGCGGGAGCGGCAGAGGATGACGCCGTTCCCT	2280
Qy	2281	GCCCAAGAGCTGAGGAACAAT	GTGCATCTCTGTCTTTGGTCTGTGGCGCTGGCGCTGTCTTA	2340





Db 461 AGGAAGTGAAGGAAGCATCAAGGGATAGTTTCAAAGTGGCGCACCGAGCGGGCC 520  
QY 722 TTGCTCTGAAATACCTTTCTCACAGAGGGTTGCTTGAGGACAGAAATGCTTCTGTGCCCC 781  
Db 521 TAGCCCTGAAA---CGCCTCAGCAGAGGGTTCCCGGAGGACAGAAATGGCTCTGTGCCCC 577  
QY 782 AGATCTCAATCATGTCATGATGGAAGTCCACAGGGGATGAGGCACTGCCATCAAGC 841  
Db 578 AGATTCTTATCATCTGTCACGAGTGGAAAGTCCACAGGGGCGGTGGCTCTCCCGCTAAGC 637  
QY 842 AGCTGAAGGAAGGGGTGTCACGTGTTTCTGCTGGGGTCAAGTTCACAGTGGGAGG 901  
Db 638 AGCTGAGAGAAAGGGGATCGTGTGTTTCCGTAGAGTCCGTTTTCAGGTGGGAGC 697  
QY 902 AGCTGATGCACATGGCCAGGACCTAGAGGGCAGCAGTGTCTGTTGGCTGAGCAGGTGG 961  
Db 698 AGCTGCTCACGCTGGCCAGTGAAGCCGAGGACCGCATGTGCTTGGCTGACGAAGTGG 757  
QY 962 AGGATGCCACCAAGGGCTTTCAGCACCTCAGCAGCTCGGCCATCTGCTCAGCGCCA 1021  
Db 758 AGGATGCCACCAATGGCTCTCAGCACCTCAGCAGCTCGGCCATCTGCAACACTGCTG 817  
QY 1022 CGCCAGACTCAGGTCGAGGCTCACCCCTGTCAGCACAGGACGCTGGAGATGTCGGG 1081  
Db 818 ATCCAGACTCGAGGTGGAACTCATCCCTGTGAGCGGAGGACGCTGGAGACCGTCAGGG 877  
QY 1082 AGTTGCTGGCAATGCCCATGCTGGAGAGATCGGGCGGACCCCTTGGCGTCTGGCTG 1141  
Db 878 AGCTGCTGGCAATGCTTGTGCTGGAGAGATCAAGGCAAGCAGACACTGTGCTGGCTC 937  
QY 1142 CACACTGTCTCTACAGCTGGAAGAGATGTTCTTAAGCCACCTGCCACCTGTCTACA 1201  
Db 938 TGCCCTGTCTCTTACAGCTGGAAGAGATGTTCCAGACACACCCCTGCCAACTGTCTA 997  
QY 1202 GGACCACTCGCCAGGCTCTGACTCCGAGCCCTGCCAGATGGAAGGACACATGTGTT 1261  
Db 998 GAACCACTCTCCAGGCTCTGACTCCAGCCCTGCCAAATGAGGACACCTGTGATTC 1057  
QY 1262 CAGAAGACTTGGAGGCTTACAGTGGCTCTGCCCCGTGGCTTGGAGGGAGGCTAACT 1321  
Db 1058 CAGAAGTGTGGATAGTACCACCTGTCTGCTGCCACTGGATTCGGAGGGAGTCACT 1117  
QY 1322 GTCCCTGAACTGAGCTGGAATGAGGCTGACCTCTCTTCCGCTGAGGACGCTCTG 1381  
Db 1118 GTCCCGGAGCTGAGCTGGAATGAGATCGATGCTCTCTCTCTCTGAGCAGTCTG 1177  
QY 1382 CGGSCACCACTCTGGAGGCTCTCTCGGGCCAAAGTCTCTGTAAGCGGTTTGTGGGG 1441  
Db 1178 CAGGCAACCAATTTGGGGGCTTCCGAGGGCCAAAGCCCTTGTCAAGCGCTTTGTGAGG 1237  
QY 1442 CCCTGCTGAGCGAGGACTCTCGGGCCGAGTGGTGTGGCCACATACAGCAGGAGCTGC 1501  
Db 1238 CCCTGCTGAGGAGGACTCCGAGCCGCGCTTGGATAGCCAGTTATGGCAGGAATCTAA 1297  
QY 1502 TGTGCGGGTG-CCTGTGGGGAGTACCAGGA-TGTGCTGACCTGGTCTGGAGCCCTCGA 1559  
Db 1298 TGTGCGGGTGCTCTCGGGGAGTACCAGCAATGTGCGGCACTGATCAGGAGCCCTGA 1357  
QY 1560 TGGCAATTCCTTCGCTGGGGCCCACTCTGACGGGAGTGGCTTCGGCAGCGCGCAGA 1619  
Db 1358 CAGCATTCCTTCAGCGGTGGCCGACCTTAACCGGAGTGGCTTGTCCAGGTGGCAGA 1417  
QY 1620 GCGTGGCTTCGGAGGCGCCACAGGACAGCCAGCCAGCTAGAGTGGTGGTTTTT 1679  
Db 1418 GCAGGCTTTGGAGTGGCCAGCAGGACTGTCAGGACAGCCAGCAGAGTAGTACTTCT 1477  
QY 1680 GCTCACTGAGTCACTCCGAGGATGAGTGTGGGGCCCGCAGCGCTGACGCAAGGCGCG 1739  
Db 1478 GCTCACTGAATCAGCTCCAGGATGAGTGTCTGGGGCCAGCAGCTCAGCAGGAGCTCG 1537  
QY 1740 AGAGCTGCTCTCTGCTAGGCTAGGCGCTGAGGCGCTGCGGGCAGAGCTGGAGAGATCAC 1799

Db 1538 GGAGCTACTCTCTCTGGCGCTGGGCAAGTGAAGTCTTGCAGGCGGAGCTGGTGAAGATCAC 1597  
QY 1800 AGCAGCCCAAAACATCTGATGTCTACTCTCGGATCTCTCAGGATCTCTTCAACCAATCCC 1859  
Db 1598 CGGTAGCCCGAAGCATGTGATGTCTCCACAGACACCTCAGGACCTGT--CAGCCAAATCC 1655  
QY 1860 TGAGCTGCGAGGGGAAGCTGTGCAAGCGGAGCGGCGGAGGTCGCGGACACAGCCCTGGA 1919  
Db 1656 AGAGCTGCAAGAGGAGCTATGCAGCCAGGACGCGGAGGCTGCCAGGACAGTCACTGGA 1715  
QY 1920 CTTCTGCTTTCATCTGTGACACCTCTGCTCAGTAGGCGCGAGAAATTTTCTCAGATGCA 1979  
Db 1716 CTTGGTCTCTCTG-TGATGCTCTG-CTCTGTGGAGCTGAGAATTTTCCCAATGCA 1773  
QY 1980 GAGCTTTGTGAGAAAGTGTGCTCCAGTTTGTAGGTGAACCTGTGAGTGCACAGGTCTCG 2039  
Db 1774 GAGCTTTCATCAGCAAAATGCACCCCTCCGGTTTGTATGTGAATCCTGTGATGACACAAGTTGG 1833  
QY 2040 CCTG 2043  
Db 1834 CCTG 1837  
RESULT 4  
US-09-984-245-64  
; Sequence 64, Application US/09984245  
; Patent No. US20020165374A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: P2004P1  
; CURRENT APPLICATION NUMBER: US/09/984,245  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/154,707  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: PCT/US98/05311  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: US 60/041,277  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/042,344  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/041,276  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/041,281  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/048,094  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,350  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,188  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,135  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/050,937  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,187  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,099  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,352  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,186  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,069  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,095  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,131  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,096  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,355  
; PRIOR FILING DATE: 1997-05-30





Db 741 TCTCTTCTATGCCGA 755

## RESULT 7

US-10-143-090-64

; Sequence 64, Application US/10143090

; Publication No. US20030069406A1

; GENERAL INFORMATION:

; APPLICANT: Young et al.

; TITLE OF INVENTION: 87 Human Secreted Proteins

; FILE REFERENCE: P2004P1

; CURRENT APPLICATION NUMBER: US/10/143,090

; CURRENT FILING DATE: 2002-05-13

; PRIOR APPLICATION NUMBER: 09/154,707

; PRIOR FILING DATE: 1998-09-17

; NUMBER OF SEQ ID NOS: 343

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 64

; LENGTH: 1668

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1664)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-143-090-64

## Query Match

Best Local Similarity 2.2%; Score 74.6; DB 9; Length 1668;

Mismatches 242; Conservative 0; Mismatches 239; Indels 14; Gaps 1;

QY 473 CAGTGGACATCATGTTCTGTAGTGGTCTTAACAGCGTCGGGAAAGGAGCTTTGAAA 532

Db 261 CAATTGACCTGGTCTTGTGTGATCGATGATCCAAAGAGTCTTGAGAGAGAGAAATTTTGAGG 320

QY 533 GGTCCAAAGCACTTTGCCATCAGCTCTGTACGGTCTGACATCAGCCCGGAGAGGCTCA 592

Db 321 TCGTGAAGCAGTTGTGCTGGAATTTATAGATTCCTTGGAAATTTCCCTCCAAAGCGGCTC 380

QY 593 GAGTGGGAGCATTCACAGTTCCAGTTCACCTCATCTCGAATTCCTCCCTTTGGATTTCATTTT 652

Db 381 GAGTGGGCTGCTCCAGTATTCACACAGGTCCACACAGATTCACCTCTGAGAACTCA 440

QY 653 CAACCCAAAGAGTGAAGGCAAGAAATCAAGAGGATGTTTTCAAAGAGGCGCACGG 712

Db 441 ACTCAGCCAAAGACATGAAAAAGCGTGGCCACATGAATACATGGAAGGGGCTCTA 500

QY 713 AGACGGAACCTGCTCTGAATACCTTCTGCACAGAGGGTTGCTGGAGGAGCAAAATGCTT 772

Db 501 TGACTGGCTGGCCCTGAAACACATGTTTGAGAGAAAGTTTACCCAGGAGAGAGGGGCCA 560

QY 773 CT-----GTGCCCCAGATCCTCATCATCTGTCATCGTCACTGATGGGAAGTCCCGAGG 818

Db 561 GGCCCTTTCCAAAGGTCGCCAGAGCAGCCATGTTTCCACGAGGAGCGGCTCAGCA 620

QY 819 GGNATGGCACTGCCATCAACGACGTGAAGAAAGGAGTGTACATGTTTCTGCTGGG 878

Db 621 TGACGCTCCGAGTGGGCGAGTAAAGCCAAAGCCCAATGATCACTATGATGCTGTTGG 680

QY 879 GGTGAGTTTCCAGGTGGGAGGAGTGTGATGACATGCGCCAGCGAGCCTAGAGGGCAGCA 938

Db 681 GGTAGAAAAGCCATTGAGAGGAAGTAACAGAGATTCCTCTGAGCCCCAACAAAGCA 740

QY 939 CGTGTGTTGGCTGA 953

Db 741 TCTCTTCTATGCCGA 755

## RESULT 8

US-10-000-512-9

; Sequence 9, Application US/10000512

; Patent No. US2002016469A1

; GENERAL INFORMATION:

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferraro, Napoleone

; APPLICANT: Genentech, Inc.

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

;

; APPLICANT: Shimkets, Richard A

; APPLICANT: Fernandes, Elma

; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

; FILE REFERENCE: 15966-556

; CURRENT APPLICATION NUMBER: US/10/000,512

; CURRENT FILING DATE: 2001-10-23

; PRIOR APPLICATION NUMBER: 09/619,252

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 60/167,785

; PRIOR FILING DATE: 1999-11-29

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 3447

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (55)...(2931)

US-10-000-512-9

## Query Match

Best Local Similarity 2.2%; Score 73.6; DB 9; Length 3447;

Mismatches 242; Conservative 0; Mismatches 239; Indels 15; Gaps 1;

QY 473 CAGTGGACATCATGTTCTGTAGTGGTCTTAACAGCGTCGGGAAAGGAGCTTTGAAA 532

Db 2021 CAATTGACCTGGTCTTGTGTGATCGATGATCCAAAGAGTCTTGAGAGAGAGAAATTTGAGG 2080

QY 533 GGTCCAAAGCACTTTGCCATCAGCTCTGTACGGTCTGACATCAGCCCGGAGAGGCTCA 592

Db 2081 TCGTGAAGCAGTTGTGCTGGAATTTATAGATTCCTTGGAAATTTCCCTCCAAAGCGGCTC 2140

QY 593 GAGTGGGAGCATTCACAGTTTCAGTTCACCTCATCTGGAATTCCTCTGGATTTCATTTT 652

Db 2141 GAGTGGGCTGCTCCAGTATTCACACAGTTCACACAGATTCACCTGAGAAACTTCA 2200

QY 653 CAACCCAAAGAGTGAAGGCAAGAAATCAAGAGGATGTTTTCAAAGAGGCGCACGG 712

Db 2201 ACTCAGCCAAAGACATGAAAAAGCGTGGCCACATGAATACATGGGAAAGCGCTCTA 2260

QY 713 AGACGGAACCTGCTCTGAATACCTTCTGCACAGAGGGTTGCTGGAGGCAAAATG--- 769

Db 2261 TGACTGGGCTGGCCCTGAAACACATGTTTGAGAGAAAGTTTACCCAAAGGGAAGGGGCCA 2320

QY 770 -----CTTCTGTGCCCCAGATCCTCATCATCTGCTCACTGATGGGAAGTCCAGG 817

Db 2321 GGCCCTTTTCCACAGAGGTGCCCAGAGAGCCATTGTTTACCAGACGACGGGCTCAGG 2380

QY 818 GGGATGTGGCACTGCCATCCAAAGAGCTGAAGAAAGGGGTGTCACTGTGTTGCTGTGG 877

Db 2381 ATGACGCTCTCCGAGTGGGCGAGTAAAGCCAAAGCCCAATGGTATCATGATGCTGTG 2440

QY 878 GGTGTCAGTTTCCAGTGGGAGGAGTGCATCCACTGCGCAGGAGCCTAGAGGGCAGC 937

Db 2441 GGGTAGAAAAGCCATTGAGAGGAAGTAACAAGAGATTGCTCTGAGCCCCAACAAAGC 2500

QY 938 AGCTGCTGTTGGCTGA 953

Db 2501 ATCTCTTCTATGCCGA 2516

## RESULT 9

US-09-905-291A-33

; Sequence 33, Application US/09905291A

; Patent No. US2002016037A1

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferraro, Napoleone





: APPLICANT: Pan, James  
 : APPLICANT: Paoni, Nicholas F.  
 : APPLICANT: Roy, Margaret Ann  
 : APPLICANT: Stewart, Timothy A.  
 : APPLICANT: Tumas, Daniel  
 : APPLICANT: Williams, P. Mickey  
 : APPLICANT: Wood, William I.  
 : TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

```

/ TITLE OF INVENTION: ACIDS Encoding the Same
/ FILE REFERENCE: 10466-14
/ CURRENT APPLICATION NUMBER: US/09/907,841
/ CURRENT FILING DATE: 2001-11-20
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 33
/ LENGTH: 3449
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-907-841-33

Query Match          2.2%; Score 73.6; DB 9; Length 3449;
Best Local Similarity 48.8%; Pred. No. 2.8e-11;
Matches 242; Conservative 0; Mismatches 239; Indels 15; Gaps

```

RESULT 12

2098	CGGAGGACAGTTGTGTCATCGGGAATATAGAGTATCTTGTGACAAATTTCCCAAGAGCCGCTC	41
QY	GACTGGGAGCATTTCCAGTTCAGTTCCACTCCTCATCTGGAATTCCTCCCTTGGATTCATTTT	652
2158	GAGTGGGGTGTCCAGTATTTCCACACAGGTCCACACAGAGTTCACTCTGAGAAACTTCA	221
Db		
QY	CACCCACACAGGAAGTGAAGGCNAGAAATCAAGAGGATGTTTTCAAGAGAGGGCCGACGG	712
2218	ACTGACCCAAAGACATGAAAAGCGTGGCCCCACATGAATACATGGGAAAAGGGCTCTA	227
Db		
QY	AGACGGAACCTTGCTCTGAAATACCTTCTGCACAGAGGGTTGCTGGAGGCACGAATG---	769
2278	TGACTGGCTTGGCCCTGAAACACATGTTTGAGAGAAGTTTTACCCNAGGAGGAAGGGGCCA	233
Db		
QY	-----CTTCTGTGCCCGAGATCCTCATCATCTGCTACTGATGGGAAGTCCCAGG	817
2338	GGCCCTTTCCACAAGGGTGCCACAGACAGCCATTGTGTCCACGACGACGCGGTCTAGG	239
Db		
QY	GGATGTGGCACTGCCATCCAAGCAGCTCAAGGAAGGGGTGCTACTGTGTTTGCTGTGG	877
2398	ATGACGTCTCCGAGTGGGGCAATGTTATCATCTATGTATGTATGTGTGTG	245
Db		
QY	GGGTCAAGTTTCCCAAGGTGGGAGGAGCTGCATGCTACTGTGCCACGAGCGCTTAGAGGSCAG	937

Db 2458 GGGTAGGAAACCCATTGAGGAGGACTACAAGAGATTGCTTCTGAGCCCAACAAGC 2517  
Qy 938 ACCTGCTGTTGGCTGA 953  
Db 2518 ATCTCTTCTATGCCGA 2533

RESULT 13

US-09-904-011-33  
; Sequence 33, Application US/09904011  
; Publication No. US20030003530A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/904,011  
; PRIOR FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 09/665,350  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 33  
; LENGTH: 3449  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-904-011-33

Query Match 2.2%; Score 73.6; DB 9; Length 3449;  
Best Local Similarity 48.8%; Pred. No. 2.8e-11;  
Matches 242; Conservative 0; Mismatches 239; Indels 15; Gaps 1;  
Qy 473 CAGTGGACATCATGTTCTCTGTAGATGGTCTTAACAGCGTCGGGAAAGGAGCGCTTGA 532  
Db 2038 CAATTGACCTGCTCTTTGTGATCGATGGAATCCCAAGAGTTCTGGAGAAGAAATTTTGG 2097  
Qy 533 GGTCAAGCACATTGGCATCACAGTCTGTGACGGTCTGCACATCAGCCCGAGAGGTCA 592  
Db 2098 TCGTGAAGCAGTTTGTCACTGGAATATAGATTCTTGACAATTTCCCCCAAGCGCTC 2157  
Qy 593 GAGTGGGACATTCAGTTTCACCTCCATCTCCATCTGGAATTTCCCTTTGGATTCTATTTT 652  
Db 2158 GAGTGGGGCTGCTCCAGTATTCACACACAGGTCCACACAGATTCTCTGAGAACTTCA 2217  
Qy 653 CAACCCACAGGAGTGAAGCAAGAAATCAAGAGGATGCTTTTCAAGAGGAGCGGCACGG 712  
Db 2218 ACTCAGCCAAAGACATGAAAAAGCGGTGGCCACATGAATACATGGAAGGGGCTCTA 2277  
Qy 713 AGACGGAACTTGTCTGAAATACCTTCTCACAGAGGGTGTGCTGAGCGCAGAAATG--- 769  
Db 2278 TCAGTGGGCTGGCCCTGAAACACATGTTTGAGAGAAGTTTACCACAGGAGAGGGGCCA 2337  
Qy 770 -----CTTCTGTGCCCCAGATCCTCATCATCTACTGATGGAAGTCCAGG 817  
Db 2338 GGCCCTTTTCCCAAGAGGTGCCACAGAGCAGCATTTGTTCACCGACGGTCCAGG 2397  
Qy 818 GGGATGTGGCACTGCCATCAAGCAGCTCAAGGAAAGGGGTGCTACTGTGTTGTGCTGG 877  
Db 2398 ATGAGCTCTCCAGTGGGCCAGTAAGCCAGGCCAATGGTATCACTATGATCTGTG 2457  
Qy 878 GGGTCAGGTTTCCAGTGGGAGGAGCTGCATGCATGCGCCAGCGAGCCTAGAGGCGACG 937  
Db 2458 GGGTAGGAAAGCATTGAGGAGGAACTACAAGAGATTGCTCTGTAGCCCAACAAGC 2517  
Qy 938 ACCTGCTGTTGGCTGA 953  
Db 2518 ATCTCTTCTATGCCGA 2533

RESULT 14

US-10-028-072-293  
; Sequence 293, Application US/10028072  
; Publication No. US20030004311A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang  
TITLE OF INVENTION:  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/028,072  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059836  
PRIOR FILING DATE: 1997-09-24  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062285  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062287  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062814  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/062816  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063045  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063082  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/063127  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063327  
PRIOR FILING DATE: 1997-10-27  
PRIOR APPLICATION NUMBER: 60/063329  
PRIOR FILING DATE: 1997-10-27  
PRIOR APPLICATION NUMBER: 60/063550  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063561  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063704  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063733  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063735  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063738  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063755  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064248  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/064809  
PRIOR FILING DATE: 1997-11-07  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065846  
PRIOR FILING DATE: 1997-11-17  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066453  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066511  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069212  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069278  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069334  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069694  
PRIOR FILING DATE: 1997-12-16  
PRIOR APPLICATION NUMBER: 60/072320  
PRIOR FILING DATE: 1998-01-23  
PRIOR APPLICATION NUMBER: 60/073612  
PRIOR FILING DATE: 1998-02-04  
PRIOR APPLICATION NUMBER: 60/074086  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/074092  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-02-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081695  
PRIOR FILING DATE: 1998-04-14  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081818  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082999  
PRIOR FILING DATE: 1998-04-24  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085149  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086414  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086430



; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 33  
; LENGTH: 3449  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-906-742-33

Query Match 2.28; Score 73.6; DB 9; Length 3449;  
Best local Similarity 48.8%; Pred. No. 2.8e-11;  
Matches 242; Conservative 0; Mismatches 239; Indels 15; Gaps 1;  
Qy 473 CAGTGGACATCATCTTCTGTAGATGGCTCAACAGCGTGGGAAAGGAGCTTTGAAA 532  
Db 2038 CAATTGACCTGGCTTTGTGTATCGATGATCCAAAGTCTTGGAGAAGAGATTTTGAGG 2097  
Qy 533 GGTCCAAAGCACTTTGCCATCACAGTCTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCA 592  
Db 2098 TCGTGAAGCAGCTTTGTCTACTGGAATATAGATTCTTTCGACAAATTTCCCAAGCCGCTC 2157  
Qy 593 GAGTGGGAGCATTCAGTTCCACTCTCATCTGGAATTCCTCCCTTGGATTTCATTTT 652  
Db 2158 GAGTGGGGCTGCTCCAGTATTTCCACAGAGTCCACAGAGTTTCACTCTGAGAAACTTCA 2217  
Qy 653 CAACCCAAAGAGAGTGAAGGCAAGATCAAGAGGATGGTTTTCAAAAGGGGCGCACGG 712  
Db 2218 ACTCAGCCAAAGACATGAAGAAAGCCGTGGCCACATGAANTACATGGGAAGGGCTCTA 2277  
Qy 713 AGACGGAACCTTGCTCTGAAATACCTTCTGCACAGAGGGTTGCCCTGGAGGAGAGAAATG--- 769  
Db 2278 TGACTGGGCTGGCCCTGAAACACATGTTTGAGAGAAGTTTTACCCAAAGGAGAAGGGGCCA 2337  
Qy 770 -----CTTCTGTGCCCCAGATCCTCATCATCGTCACTGATGGGAGTCCCAGG 817  
Db 2338 GGCCCCCTTTCACAAAGGTGGCCACAGACAGCATTTGTTTACCGAGCGAGGGGCTCAGG 2397  
Qy 818 GGGATGTGGCACTGCCATCCAGCAGCTGAAGGAAAGGGGTGTCACCTGTGTGTTGCTGTGG 877  
Db 2398 ATGAGCTCTCCGAGTGGGCCAGTAAGCCAAAGGCCAATGGTATCATCTATGTCTGTG 2457  
Qy 878 GGGTCAGGTTTCCAGGTGGGAGAGTGCATGCACTGGCCAGCGAGCCCTAGAGGGCAGC 937  
Db 2458 GGGTAGGAAAAGCCATTGAGGAGAACTACAAGAGATTGCCCTCTGAGCCCAACAAAGC 2517  
Qy 938 ACGTCTCTTGGCTGA 953  
Db 2518 ATCTCTTCTATGCCGA 2533

Search completed: May 5, 2003, 04:59:36  
Job time : 346 secs

7

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 01:38:18 ; Search time 3008 Seconds  
(without alignments)  
18171.465 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgttcggtgcac.....tgtccaccttgaagtcttc 3375

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_nam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	458.8	13.6	1024	12	BF164945	BF164945 PM4-EN0006
C	2	428.6	12.7	535	12	BF849816 PM4-EN0006
3	393	11.6	664	10	BB617396	BB617396 BB617396
C	4	389.8	11.5	632	17	AZ344444 AZ344444
5	349.4	10.4	638	10	BB613951	BB613951 BB613951
C	6	340	10.1	406	10	AW206492 UI-H-B11-1

7	303.8	9.0.	647	10	BB655280	BB655280	BB655280
8	294.8	8.7	631	10	BB612001	BB612001	BB612001
9	264.2	7.8	420	10	BB846685	BB846685	BB846685
10	219	6.5	393	12	BF606846	BF606846	BF606846
C	11	208.6	6.2	282	10	BE143292	MR0-HT016
12	176.4	5.2	629	13	BJ003847	BJ003847	BJ003847
C	13	125.2	3.7	201	10	AW856235	KC1-CT028
14	114.8	3.4	438	13	BJ036168	BJ036168	BJ036168
C	15	113	3.3	723	14	BM702190	UI-E-CQ1-
16	98.8	2.9	591	10	AV593500	AV593500	AV593500
C	17	92.4	2.7	483	17	AZ859528	2M0165K08
18	88.2	2.6	1011	14	BQ34713	BQ34713	AGENCOURT
19	88.2	2.6	1393	11	AK014342	AK014342	Mus_muscu
20	85.4	2.5	615	13	BJ499393	BJ499393	BJ499393
21	83.4	2.5	567	10	AW855319	PM1-CT026	AW855319
22	83.2	2.5	402	13	BJ067406	BJ067406	BJ067406
23	75.4	2.2	563	9	AI385774	mJ39C09_Y	AI385774
24	75.2	2.2	514	13	BI067543	pgfln_pk0	BI067543
25	74.4	2.2	623	13	BJ501694	BJ501694	BJ501694
26	73.4	2.2	510	13	BI337438	360071_MA	BI337438
27	72.8	2.2	742	13	BM050442	603632704	BM050442
28	72.2	2.1	600	10	BB577711	BB577711	BB577711
29	72.2	2.1	716	13	BI658919	603301478	BI658919
30	72.2	2.1	1149	13	BM454041	AGENCOURT	BM454041
31	72.2	2.1	1407	11	AK017553	Mus_muscu	AK017553
32	72	2.1	867	14	BQ876319	AGENCOURT	BQ876319
33	71.4	2.1	580	10	AV594887	AV594887	AV594887
34	71.4	2.1	603	10	AV592803	AV592803	AV592803
35	71.2	2.1	654	10	BB617384	BB617384	BB617384
36	71	2.1	558	10	AV599656	AV599656	AV599656
37	70.6	2.1	568	10	AV607535	AV607535	AV607535
38	70.6	2.1	590	10	BB616450	BB616450	BB616450
39	70.6	2.1	798	13	BI546662	603191404	BI546662
40	70.2	2.1	567	10	AV595999	AV595999	AV595999
41	69.8	2.1	496	17	AQ202422	RPC111-58	AQ202422
42	69.6	2.1	966	9	AL548954	AL548954	AL548954
C	43	69	536	10	BE048895	hr55c09_x	BE048895
44	69	2.0	585	13	BM313579	Ig72b01_Y	BM313579
45	69	2.0	618	14	BQ778029	I140a12_Y	BQ778029

ALIGNMENTS

RESULT 1	BF164945	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
LOCUS	BF164945	1024 bp	mus musculus	cdna	linear	EST 30-OCT-2000
DEFINITION	BF164945	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
ACCESSION	BF164945	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
VERSION	BF164945	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
KEYWORDS	EST.	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
SOURCE	EST.	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
ORGANISM	EST.	601778179F1 NCI_CGAP_Lu29	Mus musculus	cdna	linear	EST 30-OCT-2000
REFERENCE	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000
AUTHORS	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000
TITLE	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000
JOURNAL	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000
COMMENT	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000
FEATURES	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000
source	1	(bases 1 to 1024)	Mus musculus	cdna	linear	EST 30-OCT-2000

```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="IMAGE:4019899"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/site="Organ: lung; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 243 a 299 c 330 g 151 t
ORIGIN
Query Match 13.6%; Score 458.8; DB 12; Length 1024;
Best Local Similarity 79.0%; Pred. No. 2.6e-100;
Matches 571; Conservative 0; Mismatches 148; Indels 4; Gaps 2;
QY 513 CGGAAAGGAGCTTGAAGGTCACAGCACTTGGCCATCAGCTGTGACGGTCTGGA 572
Db 34 CGGAGAGGGAGCTCGAGAGTCCAAAGCGCTTGCCATCGCTGTGATGCCCTGGA 93
QY 573 CATCAGCCCGAGAGGTCAGAGTGGAGCATTCAGTTTCAGTTCACCTCCTCATCTGA 632
Db 94 CATCAGCCCTGGCAGGTCAGAGTGGAGCTTGCAGTTTGTTCACCTCCTCGTCTGA 153
QY 633 ATTCCCTTGGATTTCATTTTAAACCCACAGGAAGTGAAGGAAAGAAATCAAGAGGATGT 692
Db 154 ATTCCCTTGGACTCTTCTCAACGCGACAGAGTGAAGGAAGCATCAAGGGATATCT 213
QY 693 TTTCAAGAGGGGCGACGAGGAGCACTTGTCTGTAATACCTTCTCAGCAGAGGTT 752
Db 214 TTTCAAGGTGGCGCAGCAGAGCGGCTAGCCCTGAAA---CGCTGAGCAGAGGTT 270
QY 753 GCTGTGAGCAGAAATGCTTCTGTGCCCCAGATCTCATCTGCTACGTCAGGGAAGTC 812
Db 271 CCCCAGGAGCAGAAATGCTTGTGCCCCAGATCTTATCATCTGTCAGGATGGCAGTC 330
QY 813 CCAGGGGATGTGGCACTGCCATCCCAAGCAGCTGAAGGAAAGGGTGTCACTGTCTTGC 872
Db 331 CCAGGGGCGCTGGCTCTCCCGCTAAGCAGCTGAGAGAAAGGGGCATCGTGTGTTGC 390
QY 873 TGTGGGTCAGGTTTCCAGGTGGAGAGCTGCATGCTGCGCCAGCAGGCTAGAGG 932
Db 391 CGTAGAGTCCGTTTCCAGGTGGAGAGCTGCTTCACGCTGGCCAGTGAAGGA 450
QY 933 CGACGACGTGCTTGGCTGAGCAGGTGGAGGATGCCAACCGGCTCTTCAACACCT 992
Db 451 CCAGCATGTGCTTGGCTGAGCAAGTGGAGGATGCCACCATGGCTCTCAGCACCT 510
QY 993 CAGCAGCTCGGCATCTGTCCAGGCCACGCCAGACTGCAGGGTGGAGGCTACCCCTG 1052
Db 511 CAGCAGCTCGGCATCTGCACCACTGCTGATCCAGACTGCAGGTGGAACTCATCCCTG 570
QY 1053 TGAGCAGAGGCTGGAGAGTGGTCCGGAGTTCGCTGCAATGCCCATGCTGGAGAG 1112
Db 571 TGAGCGGAGGCTGGAGAGCTGAGGAGTCGCTGCAATGGCTTGTGTGGAGAG 630
QY 1113 ATCGCGCGGAGCCTTGGCTGCTGCTGCACACTGCTCCTCTTACAGCTGGAAGAGCT 1172
Db 631 ATCAAGGCAAGCAGACACTGTCTGGCT-CTGCTGTCTGCTTACAGCTGGAAGAGG 689
QY 1173 GTTCTTAACCCACCTTGCACCTGTCTACAGGACCACTGCTGCCAGGCCCTGTGACTCGCA 1232
Db 690 GTCCAGAAACACCTTGGCGAAATGGGACAGAACCACTCTGCGAGGCCGCTGATCCAGGCT 749
QY 1233 GCC 1235
Db 750 GGC 752
RESULT 2
```

```
BF849816 535 bp mRNA linear EST 16-JAN-2001
PM4-EN0068-151100-004-b06 EN0068 Homo sapiens cDNA, mRNA sequence.
BF849816
EST. BF849816.1 GI:12236966
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 535)
Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,
Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,
Goldman G.H., Carvalho A.F., Matsukuma A., Balg G.S., Simpson D.H.,
Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare
M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
Simpson A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM4&t2=PM4-EN0068-
151100-004-b06&t3=2000-11-15&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 534.
FEATURES
Location/Qualifiers
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0068"
/dev_stage="Adult"
/note="Organ: lung normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 142 a 150 c 124 g 119 t
ORIGIN
Query Match 12.7%; Score 428.6; DB 12; Length 535;
Best Local Similarity 97.5%; Pred. No. 3.9e-93;
Matches 467; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
QY 2899 CCCACTTCTACCTGCTGTGCTGTGAGGCTATGTCATCGCCACCTTCCCTTGA 2958
Db 535 CCCACTTCTGCTACCTGCTGTGCTGTGAGGCTATGTCATCGCCACCTTCCCTTGA 476
QY 2959 GGATAAACAAGGGTCTCAAGACTTAAATTTAGCGGCTGACGTTCCTTTGCACACAAT 3018
Db 475 GGATAAACAAGGGTCTCAAGACTTAAATTTAGCGGCTGACGTTCCTTTGCACACAAT 416
QY 3019 CAATGCTCGCCAGAATGTTGTTGACACAGTAATGCCCCAGAGGCTTTTACTAGAGCAT 3078
Db 415 CAATGCTCGCCAGAATGTTGTTGACACAGTAATGCCCCAGAGGCTTTTACTAGAGCAT 356
QY 3079 CCTTTGGAGCGGAGGGCCACGGCCCTTCAAGATGAAGAGCAGCCTTTCCACTTCCC 3138
Db 355 CCTTTGGAGCGGAGGGCCACGGCCCTTCAAGATGAAGAGCAGCCTTTTCCACTTCCC 296
QY 3139 CAGACACATCTCTGGATGCAATTTGCAATTTGAGTGAAGGGGGCTTGAGGAGCCTTTGTGA 3198
Db 3139 CAGACACATCTCTGGATGCAATTTGCAATTTGAGTGAAGGGGGCTTGAGGAGCCTTTGTGA 3198
```



```

Db 588 AGCTGAGAGAAAGGGCATCGTGGTGGTGGCCGTAGGAGTCGGTTTCCAGAGTGGGAGG 647
QY 902 AGCTGCATCGACTGGCC 918
Db 648 AGCTGCTCAGCTGGCC 664

RESULT 4
LOCUS
DEFINITION
  632 bp DNA linear GSS 29-SEP-2000
  1M0078P24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  clone UUGC1M0078P24 F, DNA sequence.
ACCESSION
  AZ344444
VERSION
  AZ344444
KEYWORDS
  GSS.
SOURCE
  AZ344444.1 GI:10423680
  house mouse
  Mus musculus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
    1 (bases 1 to 632)
    Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
    Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
    ,M., Rose,M., Stokes,R., Tingley,A., von Niederhausern,A.
    and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0078 row: P column: 24
  Seq primer: CTTGTAAACGACGGCCAGT
  Class: plasmid ends
  High quality sequence stop: 632.
  Location/Qualifiers
    1..632
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGC1M0078P24"
      /clone_lib="Mouse 10kb plasmid UUGC1M library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
      /note="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adaptor DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of PWD42 (gil47321141gb|AF129072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adaptor mouse DNA was annealed to
      adaptor vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."
BASE COUNT 134 a 199 c 178 g 120 t
ORIGIN

Query Match 11.5%; Score 389.8; DB 17; Length 632;
Best Local Similarity 83.1%; Pred. No. 1.1e-83;

```

```

Matches 456; Conservative 0; Mismatches 92; Indels 1; Gaps 1;
QY 1894 CCAGGTCGGGACACAAAGCCCTGGACCTCGCTTCATGTTGGACACCTCTCTGCCTCAGTA 1953
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 CCAGGTCGGGACACAGTCACCTGGACCTGGTCTTCTCTGTTGGATGCTCTCCCTCTG 506
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1954 GGGCCCCGAGAAATTTGGCTCAGATCAGAGCTTTGTGAGAAAGCTGTGCCCTCCAGTTTGAG 2013
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 505 GGACGTCGAGAACTTTGGCCAAATGCAGAGCTTCATCAGGAAATGCACCTCCGGTTGAT 446
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2014 GTGACCCCTGAGTCACACAGCTGGCCCTGGTGTGTATGGCAGCAGCTGAGAGTGGC 2073
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 GTGAATCCTGATGTGACACAGGTGGCCCTGGTGGTATATGGCAGCGGGGTGCAGACTGCC 386
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2074 TTCGGGCTGGACACCAAAAC-CCACCCGGGCTGCGATGCTGCGGGGCATTAGCCAGGCCCC 2132
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 385 TTCGGGCTAGACACCCATCNCACCCGGGCTGCCGTGTTGGGGCCATGAGCAGGCTCC 326
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2133 CTACCTAGTGGGGTGGGTCTAGCCGGCACCCCTCTGCTGCACATCTATGACAAAGTGAT 2192
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 CTACCTAGGCGGTGGGTTCCTGCTGGTACAGCTTGTCTGCACATTTGAGGACAAAGTGAT 266
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2193 GACCGTCAGAGGGGTGGCCGCTGCTGCCAAAGCTGTGGTGTCTCACAGGCGG 2252
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 GACGTCCAGAGGGGTGGCCGCTGCTGCCAAAGCTGTGGTGTCTCACAGGCGG 206
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2253 GAGAGCGCAGAGGATGACAGCGTTCTCTCCAGAGAGCTGAGGAAACATCTCTCT 2312
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 GAGTGGGCGAGAAGATGACAGCTGTCCCTGCCAGAGAGCTGAGAGCAATGGCATCTCAT 146
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2313 CTTGGTCTGGCGGTGGGCGCTGCTTAAGTAGAGGCTTGCAGGAGCTTGCAGGTCCTCCG 2372
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145 CTTGGTATGAGCTGGGAGCGCTCTCAGGAGGAGCAGTACGAGGCTTTCGCGGTCCACG 86
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2373 GGATTCCTCTGATCCAGTCGGCAGCTTACCCCGAGCTGCGGTACCACAGAGAGCTGCTCAT 2432
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 CGACTCCCTGATCCATGTCGGAGCTTACACAGACCTTCGTTACCATCAGACATGCTCAT 26
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2433 TGAGTGGCT 2441
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25 CGAGTGGCT 17
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BB613951
LOCUS
DEFINITION
  BB613951 RIKEN full-length enriched, 0 day neonate head Mus
  musculus cDNA clone 483241E03 5', mRNA sequence.
ACCESSION
  BB613951
VERSION
  BB613951.1 GI:16454419
KEYWORDS
  EST.
SOURCE
  house mouse
  Mus musculus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
    1 (bases 1 to 638)
    Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
    , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
    ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
    Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
    ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
    Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
    Muramatsu,M. and Hayashizaki,Y.
    RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
    Unpublished (2001)
    Contact: Yoshihide Hayashizaki
    Laboratory for Genome Exploration Research Group, RIKEN Genomic
    Sciences Center(GSC), Yokohama Institute
    The Institute of Physical and Chemical Research (RIKEN)
    1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
    Tel: 81-45-503-9222
    Fax: 81-45-503-9216
    Email: genome-res@gsc.riken.go.jp.

```

URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
.M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
.S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
.Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa  
.K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

FEATURES  
source

Location/Qualifiers  
1. .638  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, 0 day neonate  
head"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="0 day neonate"  
/lab\_host="DH10B"  
/notes="Site:1: Sali; Site:2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAAGGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATTCGAGTTAAATTAATTAATCCCGCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC 1."  
BASE COUNT 131 a 192 c 182 g 133 t  
ORIGIN  
Query Match 10.4%; Score 349.4; DB 10; Length 638;  
Best Local Similarity 78.4%; Pred. No. 7.5e-74;  
Matches 432; Conservative 0; Mismatches 116; Indels 3; Gaps 1;  
QY 302 CGCGGCTCTCTCCGTTATATCAACATGCCCGCTTCCTGTTGCTGGAGCGGCTGTG 361  
DB 90 CGCGGCTCGCACCAGGCTCTAACAGATGCTCCACTCTGCTTACCAGCATCTACA 149  
QY 362 TTTTCTGTTTTCAGAGTGCCTCCCTCTCCAGAGTCCATGTGAAGCAAG 421  
DB 150 TGCTCTGTTTTCAGAGTGCCTCCAGCATCTCTCTTCAGGAAGTGCATGTGAACCGGG 209  
QY 422 AAACCATCGGAAGATTTCAGCTGCCAGCAAAATGATGTGCTCGGCTGCAGTGGACA 481  
DB 210 AGACCATGGGAAGATCGCTGTGCCAGCAAAATTAATGTGCTGTGCTCAGCGCGGTGACA 269

QY 482 TCATGTTTCTGTAGATGGTCTTAACAGCGTCGGGAAAGGAGCTTTTGAAGGTCCAAGC 541  
DB 270 TCCTGTTTCTGTAGATGGTCTTACAGCATCGGGAAGGGAGCTTCGAGAGGTCCAAGC 329  
QY 542 ACITTTGCCATCAGCTGTGACGGTCTGACATCAGCCCGAGAGGTCAGAGTGGAG 601  
DB 330 GCTTCCGCATCGTGGTGTGATGCCCTGGACATCAGCTGCGAGGTGAGTGGAG 389  
QY 602 CATTCCAGTTTCAGTTCCTCCTCATCTGGAATTTCCCTTTGGATTCTATTTCAACCCAAC 661  
DB 390 CTTTGCAGTTTGGTTCCTCCTCATCTGGAATTTCCCTTTGGACTCTCTTCAACTCGAC 449  
QY 662 AGGAAGTGAAGGCAAGAAATCAAGAGGATGTTTTTCAAGAGGGCGCACGAGACGGAAC 721  
DB 450 AGGAAGTGAAGGAAAGCATCAAGGGGATAGTTTTTCAAGAGTGGCGCACGAGACGGCC 509  
QY 722 TTGCTCTGAATACCTTCTCCAGAGAGGTTGCTGAGGAGCAGAAATGCTTCTGTGCCCC 781  
DB 510 TAGCCCTGAAA---CGCCTGAGCAGAGGGTTTCCCGGAGGAGCAAAATGGCTTCTGTGCCCC 566  
QY 782 AGATCTCTCATCTGCTCCTCCTGATGGGAAAGTCCAGAGGGGATGTGGCACTGCCATCCAAGC 841  
DB 567 AGATTCTTATCTGTCACGAGTGGCAAGTCCAGAGGGCGCGTGGCTCTCCCGGCTAAGC 626  
QY 842 AGCTGAAGGAA 852  
DB 627 AGCTGAGAAAA 637  
RESULT 6  
AW206492/c  
LOCUS  
DEFINITION  
IMAGE:2721001 3', mRNA sequence.  
ACCESSION  
AW206492  
VERSION  
AW206492.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 406)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-re@mail.nih.gov](mailto:cgaps-re@mail.nih.gov)  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/Db/rrp/image/image.html](http://www.bio.lnl.gov/Db/rrp/image/image.html)  
Seq primer: M13 Forward  
POLYA=Yes.

Location/Qualifiers  
1. .406  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2721001"  
/lab\_host="NCI CGAP Sub3"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site:1: Not 1; Site:2: Eco RI; The  
NCI-CGAP Sub3 library is a subtracted library derived from  
the NCI-CGAP Sub1 library, which is a subtracted library  
derived from B1. B1 constitutes a mixture of 21  
normalized or subtracted NCI-CGAP libraries: NCI-CGAP\_Co4  
, NCI-CGAP\_Pr22, NCI-CGAP\_Pr28, NCI-CGAP\_Co10,  
NCI-CGAP\_Co16, NCI-CGAP\_Kid5, NCI-CGAP\_Kid12,

```

NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLE1, NCI_CGAP_Lei2,
NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831,
1473368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
LLAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,
1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,
1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
985608-986759, 1101192-1101959, 1217928-1220615);
NCI_CGAP_Co10 pool 1 LLAM 2844-2653, 2871-2872 (IMAGE
Clonoids 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described [Bonaldo, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB=NCI_CGAP_Co10
TAG_TISSUE=colon
TAG_SEQ=AAACG"
BASE COUNT 80 a 112 c 107 g 107 t
ORIGIN

Query Match 10.1%; Score 340; DB 10; Length 406;
Best Local Similarity 90.7%; Pred. No. 1.le-71;
Matches 390; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 2482 AGCCCGTGCATGAATGAGGCGAGTCGTCCTGCAGATGGAGCTACCGCTGCAAGTGT 2541
|||||
DB 406 AGCCCGTGCATGAATGAGGCGAGTCGTCCTGCAGATGGAGCTACCGCTGCAAGTGT 347
|||||

QY 2542 CGGGATGGCTGGAGGGCCCCCAGTGCAGAAACCGTGTGAGTGAGCTCTTGTCTGTATGT 2601
|||||
DB 346 CGGGATGGCTGGAGGGCCCCCAGTGCAGAAACCGTGTGAGTGAGCTCTTGTCTGTATGT 313
|||||

QY 2602 GTGAGCCAGGATGGATTTCTTGAGAGCCGCTTGAGGCACATGGCTCCGCTGAGGAGGCG 2661
|||||
DB 312 -----GATTTCTGAGAGCCGCTTGAGGCACATGGCTCCGCTGAGGAGGCG 267
|||||

QY 2662 AGCAGCCGTACCCCTCCAGCAACTACAGAGAGGCGCTGGGCACCTGAAATGGTGCCTACC 2721
|||||
DB 266 AGCAGCCGTACCCCTCCAGCAACTACAGAGAGGCGCTGGGCACCTGAAATGGTGCCTACC 207
|||||

QY 2722 TTCTGGAATGTCTGTGCCCGAGTCTTGAATGTCTGTCTCCCGCGTGGCCAGGACCA 2781
|||||
DB 206 TTCTGGAATGTCTGTGCCCGAGTCTTGAATGTCTGTCTCCCGCGTGGCCAGGACCA 147
|||||

QY 2782 CTATTCTCACTGAGGAGGAGGTGTCCTCACTGAGCCATGCTGCTTGAAGCAACAAGAA 2841
|||||
DB 146 CTATTCTCACTGAGGAGGAGGTGTCCTCACTGAGCCATGCTGCTTGAAGCAACAAGAA 87
|||||

QY 2842 CGAGCTGATGTCAACCAACAGATGTTGTTGAAAGTTTTGATGTGAAGTAATAATACC 2901
|||||
DB 86 CGAGCTGATGTCAACCAACAGATGTTGTTGAAAGTTTTGATGTGAAGTAATAATACC 27
|||||

QY 2902 ACTTCTGTA 2911
|||||
DB 26 ACTTCTGTA 17
|||||

RESULT 7
BB655280
LOCUS
DEFINITION
BB655280 RIKEN full-length enriched, 9 days embryo Mus musculus
CDNA clone D030048B17 5', mRNA sequence.

```

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

## REFERENCE AUTHORS

## TITLE JOURNAL COMMENT

## FEATURES source

BB655280 1 GI:16489108  
EST.  
Mus musculus  
house mouse  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 647)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.  
Hiramoto,K., Hori,F., Ishi,Y., Ito,M., Kawai,J., Konno,H., Kouda  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

Location/Qualifiers  
1..647  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D030048B17"  
/dev\_stage="9 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCCGCACTCGAGTTTTTTTTTTTTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGATTCGAGTTAATTAATATCCCTCCCTCCCTCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified



Db 208 TGCTCTCTGTTCTCAGAGTGTCCCGACCATCTCTCTTCAGGAAGTGCAATGTGAACCGGG 267

QY 422 AAACATCGGGAGATTTCAGCTGCCAGCAAAATGATGTGGCTCGCTGCGTGCAGTGGACA 481

Db 268 AGACCATGGGAAGATCGCTGTGGCGCAGCAAAATTAATGTGTGCTCAGCGCGGTGCGACA 327

QY 482 TCATGTTTCTGTAGATGGTCTAACAGCTGCGGAAGGGAGCTTTGAAGGTCTCAACG 541

Db 328 TCTGTTTCTGTAGATGGTCTCAGACATCGGAAGGGAGCTTCGAGAGGTCTCAACG 387

QY 542 ACTTTCCTCATCAGTCTCTGAGGTTCTGCACATCAGCCCCGAGAGGGTTCAGAGTGGGAG 601

Db 388 GCTTGCCTCATCGCTGCTGTGATGCCCTGCACATCAGCCCTGGCAGGGTTCAGAGTGGGAG 447

QY 602 CATTCCAGTTCAGTCCACTCTCTATCTGGAATTCCTCCCTGATTCATTTTCAACCAAC 661

Db 448 CCTTGCAGTTTGTGCTCCTCATCTGGAATTCCTCCCTGATTCATTTTCAACCAAC 507

QY 662 AGGAAGTGAAGCAGCAATCAGAGGATGTTTTCNAAGAGGGCGCAGAGAGGGAAC 721

Db 508 AGGAAGTGAAGGAGCAATCAGAGGATGTTTTCNAAGAGGGCGCAGAGAGGGAAC 567

QY 722 TTGCTCTGAATACCTTCTGCACAGAGGTTGCTGAGGCGAGAAATGCTTCTGTGCC 780

Db 568 TAGCCCTGAA---CGCCTGACAGAGGGGTCGCGGAGGCGAGAAATGCTTCTGTGCC 622

RESULT 9

BB846685

LOCUS

DEFINITION

BB846685

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

Location/Qualifiers

1. 420

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="F530004C06"

/clone\_lib="RIKEN full-length enriched, adult male kidney"

/sex="male"

/tissue\_type="kidney"

/dev\_stage="adult"

/lab\_host="SOLR"

/note="Site 1: XhoI; Site 2: SstI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5']

GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3', cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5']

GAGAGAGAGAGGATCCCAAGACTCAATTAATTAATTAACCCCCCCCC 3'].

cDNA was cleaved with XhoI and SstI. "

BASE COUNT 84 a 120 c 141 g 75 t

ORIGIN

Query Match 7.8%; Score 264.2; DB 10; Length 420;

Best Local Similarity 80.3%; Pred. No. 3e-53;

Matches 335; Conservative 0; Mismatches 78; Indels 4; Gaps 2;

QY 699 AGGAGGCGCAGGAGAGGAACTTGTCTGTAATACCTTCTGCACAGAGGTGCTGG 758

Db 2 AGGTGGCGCAGGAGCGGGCTAGCCCTGAAA---CGCCTGAGCAGAGGCTTCCCGG 58

QY 759 AGGCAGAAATGCTTCTGTGCCCGCAGATCTCATCTGCTGCTGAGGAGTCCCGAGG 818

Db 59 AGGCAGAAATGCTTCTGTGCCCGCAGATCTTATCATCTGCTGAGGAGTCCCGAGG 118

QY 819 GGATGTGGCAGTCCCATCCAGAGAG---CTGAAGGAAAGGGGTGTCTGTGTGTGG 877

Db 119 GCCCGTGGCTCTCCCGCTAAGCAGCTCAGAGAAAGGGGCGATCGTGTGTGGCTAG 178

QY 878 GGGTCAGGTTTCCAGTGGGAGGAGCTGCATGCTGCCAGCGAGCTAGAGGCGAGC 937

Db 179 GAGTCGGTTCCTCCAGTGGGAGGAGCTGCTCAGCTGCGCAGTGAAGGAGGAGCGGC 238

QY 938 AGCTGCTGTGTGCTGAGCAGGTGGAGGATGCCACCAACGGCTCTTCACACCCCTCAGCA 997

Db 239 ATGTGCTGTGTGCTGAGCAAGTGGAGGATGCCACCAATGGCTCTTCACACCCCTCAGCA 298

QY 998 GCTCGGCCATCTGCTCCAGCGCCAGCAGCTGCGAGGCTCGAGGCTCACCCCTGTGAGC 1057

Db 299 GCTCCGCACTCTGAACCACTGTGATCCAGAGTGGAGGCTGGAACCTCATCTCTGTGAGC 358

QY 1058 ACAGGACGCTGGAGATGCTCGGGAGTTCGCTGGCAATGCCCATGCTCGAGAGGAT 1114

Db 359 GGAGGACGCTGGAGAACGCTCAGGAGGAGTCTGCTGGCAATGCCCTGTGCTGAGAGGAT 415

RESULT 10

BF606846

LOCUS

DEFINITION

BF606846

ACCESSION

VERSION

BF606846.1 GI:11707864

393 bp mRNA linear EST 25-APR-2001

273996 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

```
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 393)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perle, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCCGTCAGGACG
Plate: 57 row: K column: 1
Seq primer: ATTTAGTGCACATATAG.
FEATURES
source
1..393
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 64 a 130 c 123 g 76 t
ORIGIN
Query Match 6.5%; Score 219; DB 12; Length 393;
Best Local Similarity 84.5%; Pred. No. 2.8e-42;
Matches 246; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1895 CAGGGTCCGGACACAGCCCTCGCTCTTCATGTTGGACACCTCTGCCTCAGTAG 1954
Db 103 CAGGTCGCCAGGCCAGTCGCTGGACCTGCTCTTCATGTTGGACGCTCGGTCTCAGTGG 162
QY 1955 GGCCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAGCTGTGCCCTCCAGTTGAGG 2014
Db 163 GGCCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAGCTGTGCCCTCCAGTTGAGG 222
QY 2015 TGAACCTTGACGTGACACAGCTGGCTGGTGTGTATGGCAGCCAGGTGCAGCTGCCT 2074
Db 223 TGAACCCCGATGTGACGAGATCGGCTGTGGTGTGTACGGCAGCCGTGTCACAGACGCT 282
QY 2075 TCGGGCTGGACACCAACCCACCGGCTCGCATGCTGGGGCCATTAGCCAGGCCCTT 2134
Db 283 TCGGGCTGGACACCCACCTACGCGGTGCTGACGTGCTGGGGCCCTTGAGCCAGGCCCTT 342
QY 2135 ACCTAGGTGGGTGGCTCAGCCGGCAGCCGCTCGTGTGACATCTATGACA 2185
Db 343 ACCTGGAGGGGTGGCTCAGCAGGACGCGACTTGTGACATCTATGACA 393
RESULT 11
BE143292/c 282 bp mRNA linear EST 21-JUN-2000
LOCUS MR0-HT0161-221099-002-c08 HT0161 Homo sapiens cDNA, mRNA sequence.
```

```
ACCESSION BE143292
VERSION BE143292.1
KEYWORDS GI:8606013
SOURCE EST.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 282)
Dias Neto, E., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,
Nagai, M.A., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brustein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=st2-MR0-HT0161-221
099-002-c08&t3=1999-10-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 282.
FEATURES
source
1..282
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0161"
/dev_stage="Adult"
/Note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESPEC PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 71 a 78 c 63 g 70 t
ORIGIN
Query Match 6.2%; Score 208.6; DB 10; Length 282;
Best Local Similarity 93.7%; Pred. No. 7.8e-40;
Matches 239; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
QY 3026 CGCCAGAATGTTCTTCACACAGTAATGCCAGCAGAGGCCCTTTACTAGACATCCTTTGG 3085
Db 282 CGCCAGAATGTTCTTCACACAGTAATGCCAGCAGAGGCCCTTTACTAGACATCCTTTGG 223
QY 3086 ACGCGAAGGCCACGGCCCTTCAAGATGGAAGCAGCAGCTTTTCCACTTCCCAGAGAC 3145
Db 222 ACGCGAAGGCCACGGCCCTTCAAGATGGAAGCAGCAGCTTTTCCACTTCCCAGAGAC 163
QY 3146 ATTCTCG-ATGCATTTGCATTGAGTCTGAAGGGGCTTGAGGAGCTTTGTGACTTCTT 3204
Db 162 ATTCTGGAATGCAATTTGCATTGAGTCTGAAGGGGCTTGAGGAGCTTTGTGACTTCTT 104
QY 3205 GCGCACTGCTTTTGTGTGGAAGAGACTTGGAAAGGCTCTCAGACTGAATGTGACCAAT 3264
Db 103 GCGCACTGCTTTTGTGTGGAAGAGACTTGGAAAGGCTCTCAGACTGAATGTGACCAAT 44
QY 3265 TAACAGCTTGGTTG 3279
Db 43 TACCACCATGATTG 29
```

[illegible]

```
Db 74 TGCAGGTCCTCCGGGATTCC---TGACCACGTGACGTACGCCGACCTGCGAACCTATGA 18
Qy 2421 GG 2422
||
Db 17 GG 16

RESULT 14
BJ036168 438 bp mRNA linear EST 06-DEC-2001
LOCUS
DEFINITION
BJ036168 NIBB Mochii normalized xenopus neurula library xenopus
laevis cDNA clone XL018k07 5', mRNA sequence.
ACCESSION
BJ036168
VERSION
BJ036168.1 GI:17396753
KEYWORDS
EST.
SOURCE
African clawed frog.
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 438)
AUTHORS
Kitayama.A., Terasaka.C., Mochii.M., Ueno.N., Shin-i.T. and Kohara
.Y.
TITLE
Expressed genes in X. laevis embryo
JOURNAL
Unpublished (2001)
COMMENT
Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshinigenes.nig.ac.jp.
FEATURES
Location/Qualifiers
source
1..438
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_lib="NIBB Mochii normalized Xenopus neurula
library"
/tissue_type="whole embryo"
/dev_stage="stage 15"
/notes="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; CDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
```

```
BASE COUNT 106 a 112 c 101 g 113 t 6 others
ORIGIN

Query Match 3.4%; Score 114.8; DB 13; Length 438;
Best Local Similarity 60.7%; Pred. No. 6.2e-17;
Matches 182; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 338 TCCTGTGTCGGAGCGCTGCTGTTTCTCTGTTTTCCTGAGTCCGCCATCTCTCCCTC 397
|||||
Db 139 TGCCGCTGGGTTAATGGGATCTCTAGCTTCTGAGTCCCGAGTCTCAATGCTGTGTG 198
|||||
Qy 398 TCCAGGAAGTCCATGTAGCAAAACCACTCGGGAAGATTTCAGCTGCCAGCAAAATGA 457
|||||
Db 199 TCCAGGAGCTTCAGCTCAACGCAGAGACTGCCAAACAAGATTTCAGCAGAGAGACGGA 258
|||||
Qy 458 TGTGCTGCTCGGCTGACATGATGTTCTGTGTAGATGGCTCTAACAGCGCTCGGGA 517
|||||
Db 259 TGCATGTTCTTCCCACTGGACATCTCTATTCTACTGGAGCGCTCCACAGCATCGGGA 318
|||||
Qy 518 AAGGAGCTTTGAAAGCTTCCAAAGCATTTTGGCCATCACAGTCTGTGACGCTTGGACATCA 577
|||||
Db 319 GGGGTAGTTTGTAGAGGTCCTCAAGCATTTTGCCTCCAAAGTTATGTGATGCTTANATATG 378
|||||
Qy 578 GCCCGAGAGGGTCAGATGGGACCATTCAGTTCAGTTCCTCCATCTCCATCTGGAATTC 637
|||||
Db 379 GCTCTGATCTGATACNAGTGGGGGCCCATTCGAATACAGNAGTCCGCCANGNTGAGTTCC 438
|||||
```

RESULT 15  
BM702190/C  
LOCUS  
DEFINITION  
UI-E-CQ1-aey-m-08-0-UI.r1 UI-E-CQ1 Homo sapiens cDNA clone  
UI-E-CQ1-aey-m-08-0-UI 5', mRNA sequence.  
ACCESSION  
BM702190  
VERSION  
BM702190.1 GI:19015448  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 723)  
AUTHORS  
Bonaldo.M.F., Lennon.G. and Soares.M.B.  
TITLE  
Normalization and subtraction: two approaches to facilitate gene  
discovery.  
Genome Res. 6 (9), 791-806 (1996)  
CONTACT  
Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
The following repetitive elements were found in this cDNA  
sequence: 353-608. >LINE2  
Seq primer: M13 Reverse.  
Location/Qualifiers  
source  
1..723  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-CQ1-aey-m-08-0-UI"  
/clone\_lib="UI-E-CQ1"  
/tissue\_type="optic nerve"  
/dev\_stage="adult"  
/notes="Organ: eye; Vector: p7T3-pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-E-CQ1 is a normalized cDNA library containing the  
following tissue(s): optic nerve. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into p7T3-pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
CCATTAAAGTC. This library was created for the program, Gene  
Discovery in the Visual System, supported by National Eye  
Institute (NEI)."

BASE COUNT 147 a 241 c 142 g 190 t 3 others  
ORIGIN

Query Match 3.3%; Score 113; DB 14; Length 723;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2615 GGATTCTTGAGAGCCCTCGAGGCACATGCTCCCGTCGAGGAGGCGAGCGGTACCC 2674  
|||||  
Db 113 GGATTCTTGAGAGCCCTCGAGGCACATGCTCCCGTCGAGGAGGCGAGCGGTACCC 54  
|||||  
Qy 2675 CTCCAGCAACTACAGAGAGGCTGGGCGCTGAAATGGTGCCTTACCTTCTGG 2727

Db 53 CCCCCAGCACTACAGAGAGGCGCTGGCACTGAAATGGTGCGCTACCTTCGG 1

Search completed: May 5, 2003, 04:53:50  
Job time : 3030 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:52:36 ; Search time 5744 Seconds  
(without alignments)  
17099.925 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgtcgcggtgcac.....tgtcaccttgaaggtcttc 3375

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl :

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_cm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_ind.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgt\_mus.\*
- 34: em\_hgt\_pln.\*
- 35: em\_hgt\_rod.\*
- 36: em\_hgt\_nam.\*
- 37: em\_hgt\_vrt.\*
- 38: em\_sy.\*
- 39: em\_hgtgo\_hum.\*
- 40: em\_hgtgo\_mus.\*
- 41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	589	17.5	123110	9	AC005383	Homo sapi
C 2	316	9.4	160420	9	AC022023	Homo sapi
C 3	264	7.8	136357	2	AL161942	Homo sapi
4	69	2.0	496	11	G50805	SHGC-B3858
5	31	0.9	154748	2	AC12150	Mus muscu
6	25	0.7	1728	9	F243081S03	Homo sapi
7	25	0.7	11344	9	AF034611	Homo sapi
8	25	0.7	147280	9	AC067747	Homo sapi
9	25	0.7	152895	2	AC119647	Rattus no
C 10	23	0.7	622	9	HSNOTCH2	H. sapiens N
11	23	0.7	789	9	HSU77493	Human Notch
12	23	0.7	924	9	HS4296168	Homo sapi
13	23	0.7	1099	9	HS4406936	Homo sapi
14	23	0.7	9722	9	AF308601	Homo sapi
15	23	0.7	9723	6	AR102329	Sequence
16	23	0.7	9723	6	AX357059	Sequence
17	23	0.7	9723	6	AX417417	Sequence
18	23	0.7	11189	9	AF315356	Homo sapi
19	23	0.7	109612	9	AL512503	Human DNA
C 20	23	0.7	142565	10	AL592545	Mouse DNA
21	23	0.7	161987	9	AC006070	Homo sapi
22	23	0.7	174032	2	AC025904	Homo sapi
C 23	23	0.7	180836	2	AC025241	Homo sapi
C 24	23	0.7	182571	2	AC095873	Rattus no
25	23	0.7	197658	2	AL646094	Mus muscu
C 26	23	0.7	347572	2	AC107303	Homo sapi
27	22	0.7	56030	2	AC129682	Rattus no
C 28	22	0.7	90935	2	AP003917	Oryza sat
29	22	0.7	152895	2	AC119647	Rattus no
C 30	22	0.7	161987	9	AC006070	Homo sapi
C 31	22	0.7	174032	2	AC025904	Homo sapi
C 32	22	0.7	188718	2	AC113703	Rattus no
C 33	22	0.7	201802	2	AC124473	Mus muscu
C 34	22	0.7	264009	2	AC115294	Mus muscu
C 35	22	0.7	318085	2	AC127297	Mus muscu
36	21	0.6	1100	5	AF300613	Gallus ga
37	21	0.6	1182	4	AF304108	Mesoplo
38	21	0.6	1730	9	AB063081	Macaca fa
C 39	21	0.6	3024	9	AK055872	Homo sapi
C 40	21	0.6	4865	9	HSHP201	Human hp2 g
C 41	21	0.6	5421	9	CHFRAPC1	Chimpanzee
42	21	0.6	7319	10	AF164486	Rattus no
C 43	21	0.6	11551	9	HUMHPARS1	Human hapto
C 44	21	0.6	38542	9	HUMHP2HPR	Human hapto
45	21	0.6	46638	9	AC104823	Homo sapi

ALIGNMENTS

RESULT 1

AC005383

LOCUS

DEFINITION

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

AC005383 Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,  
complete sequence.

AC005383.1 GI:3818355

HTG.

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 123110)
AUTHORS	Smith,D.R.
TITLE	Direct Submission
JOURNAL	Submitted (05-AUG-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
REFERENCE	3 (bases 1 to 123110)
AUTHORS	Smith,D.R.
TITLE	Direct Submission
JOURNAL	Submitted (31-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
COMMENT	On Oct 31, 1998 this sequence version replaced gi:3808081.
FEATURES	Location/Qualifiers
source	.1.123110 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10" /map="10q25.1" /clone="Cf7987SK-1144G6"
BASE COUNT	31394 a 28668 c 29634 g 33413 t 1 others
ORIGIN	
Query Match	17.5%; Score 589; DB 9; Length 123110;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 589; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	2615 GGATTCCTTGAGACGCCCTCGAGGCACATGGCTCCCGTGCAGAGGGCAGCAGCGGTACC 2674
Db	78153 GGATTCCTTGAGACGCCCTCGAGGCACATGGCTCCCGTGCAGAGGGCAGCAGCGGTACC 78212
QY	2675 CTCCCAGCACTACAGAGAAGGCGCTGGGCACCTGAATGGTGCCTACCTTCCTGGAATGTCT 2734
Db	78213 CTCCCAGCACTACAGAGAAGGCGCTGGGCACCTGAATGGTGCCTACCTTCCTGGAATGTCT 78272
QY	2735 GTGCCCCAGGTCCCTTAGAATGTCCTGCTCCCGCGTGGCCAGGACCACACTATTCTCACTGA 2794
Db	78273 GTGCCCCAGGTCCCTTAGAATGTCCTGCTCCCGCGTGGCCAGGACCACACTATTCTCACTGA 78332
QY	2795 GGGAGGAGGATGCCCAACTGCAGCCATGCTGTTAGAGACAAGAAAGCAGCTGATGTCA 2854
Db	78333 GGGAGGAGGATGCCCAACTGCAGCCATGCTGTTAGAGACAAGAAAGCAGCTGATGTCA 78392
QY	2855 CCCAACACCATGTGTTGAAAGTTTTGATGTGTAAGTAATACCACACTTCTGTACCT 2914
Db	78393 CCCAACACCATGTGTTGAAAGTTTTGATGTGTAAGTAATACCACACTTCTGTACCT 78452
QY	2915 GCTGTGCTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAAGGGGTC 2974
Db	78453 GCTGTGCTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAAGGGGTC 78512
QY	2975 CTGAAGACTTAATTTAGCGGCGTGCAGTTCCTTTGCACACACAATCAATGCTCGCAGAAT 3034
Db	78513 CTGAAGACTTAATTTAGCGGCGTGCAGTTCCTTTGCACACACAATCAATGCTCGCAGAAT 78572
QY	3035 GTTCTTGACACAGTAATGCCAGCAGAGGCGTTTACTAGAGCATCCTTTGACGCGCGAAG 3094
Db	78573 GTTCTTGACACAGTAATGCCAGCAGAGGCGTTTACTAGAGCATCCTTTGACGCGCGAAG 78632
QY	3095 GCCACGGCCTTCCAAGATGAAAGCAGCAGCTTTTCCACTTCCCCAGAGACATTTCTGGAT 3154
Db	78633 GCCACGGCCTTCCAAGATGAAAGCAGCAGCTTTTCCACTTCCCCAGAGACATTTCTGGAT 78692
QY	3155 GCATTTGCATTTGAGTCGAAGAGGGGCGTTGAGGACGTTTGTGACTTCT 3203
Db	78693 GCATTTGCATTTGAGTCGAAGAGGGGCGTTGAGGACGTTTGTGACTTCT 78741

RESULT 2	
AC022023/c	
LOCUS	AC022023
DEFINITION	Homo sapiens chromosome 10 clone RP11-411P18, complete sequence.
ACCESSION	AC022023
VERSION	AC022023.7
	GI:20901843
	160420 bp
	DNA linear
	PRI 17-MAY-2002

KEYWORDS	HTG.	BASE COUNT	ORIGIN
SOURCE	Homo sapiens.	48686 a 32379 c 32258 g 47097 t	
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 160420)		
TITLE	Smith,D.R.		
JOURNAL	Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data		
REFERENCE	2 (bases 1 to 160420)		
AUTHORS	Smith,D.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA		
REFERENCE	3 (bases 1 to 160420)		
AUTHORS	Smith,D.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-JUN-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA		
REFERENCE	4 (bases 1 to 160420)		
AUTHORS	Smith,D.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAY-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA		
COMMENT	On May 17, 2002 this sequence version replaced gi:14522958.		
FEATURES	Location/Qualifiers		
source	1..160420		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/clone="RP11-411p18"		
	/clone_lib="RPC1-11"		
BASE COUNT	48686 a 32379 c 32258 g 47097 t		
ORIGIN			
Query Match	9.4%; Score 316; DB 9; Length 160420;		
Best Local Similarity	100.0%; Pred. No. 1.7e-172;		
Matches 316; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	2 ACAGTGTCGGCGTCACCGCTCGGAGCGTGGGTGACCGCGTAGAGTGAAGTCAACTT 61		
Db	5373 ACAGTGTCGGCGGTGCACCGCTCGGAGCTGGGTGACCGCGTAGAGTGAAGTCACTT 5314		
QY	62 TTTATTTGCAGACCTGGGCGCGATCCGCTTTAAAAAACGCGAGGGGCTCTATGCACCTCC 121		
Db	5313 TTTATTTGCAGACCTGGGCGCGATCCGCTTTAAAAAACGCGAGGGGCTCTATGCACCTCC 5254		
QY	122 CTGCGCGTAGTTCCTCGACCTCAGCGGGTCGGGTGTCGCCGCCCTCTCCAGAGAGA 181		
Db	5253 CTGCGCGTAGTTCCTCGACCTCAGCGGGTCGGGTGTCGCCGCCCTCTCCAGAGAGA 5194		
QY	182 CAACAGGTGTCCACAGTGGCAGCGCGGCCCGCCCTCTCTGTGATCCCGTAGCGC 241		
Db	5193 CAACAGGTGTCCACAGTGGCAGCGCGGCCCGCCCTCTCTGTGATCCCGTAGCGC 5134		
QY	242 CCCCTGGCCGAGCGCGCCGGGTCTGTAGTAGAGCCGCCGGGACCCAGCGCTGGT 301		
Db	5133 CCCCTGGCCGAGCGCGCCGGGTCTGTAGTAGAGCCGCCGGGACCCAGCGCTGGT 5074		
QY	302 CGCGCTCTCCTTCCG 317		
Db	5073 CGCGCTCTCCTTCCG 5058		
RESULT 3			
LOCUS	AL161942	136357 bp	DNA linear HTG 13-JUN-2001
DEFINITION	Homo sapiens chromosome 20 clone RP11-83B5, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.		
ACCESSION	AL161942		
VERSION	AL161942.5	GI:9931708	
KEYWORDS	HTG; HTGS PHASE1; HTGS CANCELLED.		

RESULT 3					
AL161942/c					
LOCUS	AL161942	136357 bp	DNA	linear	HTG 13-JUN-2001
DEFINITION	Homo sapiens chromosome 20 clone RP11-83B5,				*** SEQUENCING IN
	PROGRESS ***, 33 unordered pieces.				
ACCESSION	AL161942				
VERSION	AL161942.5	GI:9931708			
KEYWORDS	HTG; HTGS PHASE1;				HTGS CANCELLED.

```

SOURCE      human..
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 136357)
JOURNAL     Burton, J.
COMMENT     Direct Submission
            Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Aug 28, 2000 this sequence version replaced gi:9863622.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: BA8395
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 114775 bases at least Q40
            Consensus quality: 123431 bases at least Q30
            Consensus quality: 128697 bases at least Q20
            Insert size: 133157; sum-of-contigs
            Quality coverage: 2.20x in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 5425: contig of 5425 bp in length
            * 5426 5525: gap of 100 bp
            * 5526 9483: contig of 3958 bp in length
            * 9484 9583: gap of 100 bp
            * 9584 14280: contig of 4697 bp in length
            * 14281 14380: gap of 100 bp
            * 14381 17312: contig of 2932 bp in length
            * 17313 17412: gap of 100 bp
            * 17413 22848: contig of 5436 bp in length
            * 22849 22948: gap of 100 bp
            * 22949 26090: contig of 3142 bp in length
            * 26091 26190: gap of 100 bp
            * 26191 28215: contig of 2025 bp in length
            * 28216 28315: gap of 100 bp
            * 28316 33862: contig of 5547 bp in length
            * 33863 33962: gap of 100 bp
            * 33963 44145: contig of 10183 bp in length
            * 44146 44245: gap of 100 bp
            * 44246 53165: contig of 8920 bp in length
            * 53166 53265: gap of 100 bp
            * 53266 57412: contig of 4147 bp in length
            * 57413 57512: gap of 100 bp
            * 57513 60170: contig of 2658 bp in length
            * 60171 60270: gap of 100 bp
            * 60271 65985: contig of 5715 bp in length
            * 65986 66085: gap of 100 bp
            * 66086 69094: contig of 3009 bp in length
            * 69095 69194: gap of 100 bp
            * 69195 71736: contig of 2542 bp in length
            * 71737 71836: gap of 100 bp
            * 71837 74955: contig of 3119 bp in length
            * 74956 75055: gap of 100 bp
            * 75056 77325: contig of 2270 bp in length
            * 77326 77425: gap of 100 bp
            * 77426 80709: contig of 3284 bp in length
            * 80710 80809: gap of 100 bp
            * 80810 82943: contig of 2134 bp in length
            *
            * 82944 83043: gap of 100 bp
            * 83044 85218: contig of 2175 bp in length
            * 85219 85318: gap of 100 bp
            * 85319 89861: contig of 4543 bp in length
            * 89862 89961: gap of 100 bp
            * 89962 92824: contig of 2863 bp in length
            * 92825 92924: gap of 100 bp
            * 92925 95822: contig of 2898 bp in length
            * 95823 95922: gap of 100 bp
            * 95923 98130: contig of 2208 bp in length
            * 98131 98230: gap of 100 bp
            * 98231 100594: contig of 2364 bp in length
            * 100595 100694: gap of 100 bp
            * 100695 109043: contig of 8349 bp in length
            * 109044 109143: gap of 100 bp
            * 109144 112146: contig of 3003 bp in length
            * 112147 112246: gap of 100 bp
            * 112247 117183: contig of 4937 bp in length
            * 117184 117283: gap of 100 bp
            * 117284 121861: contig of 4578 bp in length
            * 121862 121961: gap of 100 bp
            * 121962 124258: contig of 2297 bp in length
            * 124259 124358: gap of 100 bp
            * 124359 127036: contig of 2678 bp in length
            * 127037 127136: gap of 100 bp
            * 127137 133357: contig of 6221 bp in length
            * 133358 133457: gap of 100 bp
            * 133458 136357: contig of 2900 bp in length.
            *
            FEATURES
            source
            1..136357
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="20"
            /clone="RP11-83B5"
            /clone.lib="RPC1-11.1"
            1..5425
            /note="assembly_fragment:01370"
            fragment_chain:1
            clone_end:SP6
            vector_side:left
            5526..9483
            /note="assembly_fragment:01120"
            fragment_chain:1
            9584..14280
            /note="assembly_fragment:00095"
            fragment_chain:1
            14381..17312
            /note="assembly_fragment:01155"
            fragment_chain:1
            17413..22848
            /note="assembly_fragment:00025"
            fragment_chain:2
            22949..26090
            /note="assembly_fragment:00387"
            fragment_chain:2
            26191..28215
            /note="assembly_fragment:00262"
            fragment_chain:3
            28316..33862
            /note="assembly_fragment:00134"
            fragment_chain:3
            33963..44145
            /note="assembly_fragment:00605"
            fragment_chain:4
            44246..53165
            /note="assembly_fragment:00989"
            fragment_chain:4
            53266..57412
            /note="assembly_fragment:00876"
            fragment_chain:5
            57513..60170
            /note="assembly_fragment:00248"
            fragment_chain:5

```

misc\_feature 60271..65985  
/note="assembly\_fragment:01554  
fragment\_chain:6"  
66086..69094  
/note="assembly\_fragment:00470  
fragment\_chain:6"  
69195..71736  
/note="assembly\_fragment:01566  
fragment\_chain:7"  
71837..74955  
/note="assembly\_fragment:00477  
fragment\_chain:7"  
75056..77325  
/note="assembly\_fragment:00036"  
77426..80709  
/note="assembly\_fragment:00390"  
80810..82943  
/note="assembly\_fragment:00409"  
83044..85218  
/note="assembly\_fragment:00437"  
85319..89861  
/note="assembly\_fragment:00764"  
89962..92824  
/note="assembly\_fragment:00782"  
92925..95822  
/note="assembly\_fragment:00808"  
95923..98130  
/note="assembly\_fragment:00831"  
98231..100594  
/note="assembly\_fragment:00886"  
100695..109043  
/note="assembly\_fragment:00992"  
109144..112146  
/note="assembly\_fragment:01007"  
112247..117183  
/note="assembly\_fragment:01102"  
117284..121861  
/note="assembly\_fragment:01146"  
121962..124258  
/note="assembly\_fragment:01320"  
124359..127036  
/note="assembly\_fragment:01489"  
127137..133357  
/note="assembly\_fragment:01498"  
133458..136357  
/note="assembly\_fragment:01604"

BASE COUNT 38044 a 27571 c 27577 g 39953 t 3212 others  
ORIGIN

Query Match 7.8%; Score 264; DB 2; Length 136357;  
Best Local Similarity 100.0%; Pred. No. 5.2e-142; Indels 0; Gaps 0;  
Matches 264; Conservative 0; Mismatches 0;  
QY 2 ACAGTGTTCGGCGCTGCACCGCTCGGAGGCTGGTGCACCGCTAGAGTGAAGTACTTT 61  
DB 9069 ACAGTGTTCGGCGCTGCACCGCTCGGAGGCTGGTGCACCGCTAGAGTGAAGTACTTT 9010  
QY 62 TTTATTGAGACCTGGCGCGATGCCGCTTTAAAAACGCGAGGGGCTCTATGCACCTCC 121  
DB 9009 TTTATTGAGACCTGGCGCGATGCCGCTTTAAAAACGCGAGGGGCTCTATGCACCTCC 8950  
QY 122 CTGGCGGTAGTCTCTCCGACCTCAGCCGCTCGGTCGTCGCCCTCTCCAGGAGAGA 181  
DB 8949 CTGGCGGTAGTCTCTCCGACCTCAGCCGCTCGGTCGTCGCCCTCTCCAGGAGAGA 8890  
QY 182 CAAACAGGTGTCCACGTGGCAGCCGCGCCGCGGCGCCCTCTCTGTATCCGTAGCGC 241  
DB 8889 CAAACAGGTGTCCACGTGGCAGCCGCGCCGCGGCGCCCTCTCTGTATCCGTAGCGC 8830  
QY 242 CCCCTGGCCGACGCGCGCCGCGG 265  
DB 8829 CCCCTGGCCGACGCGCGCGG 8806

RESULT 4  
G50805  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Michael Olivier, David R. Cox  
Stanford Human Genome Center  
Stanford University School of Medicine  
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA  
Tel: (650) 320-5800  
Fax: (650) 320-5801  
Email: olivier@shgc.stanford.edu  
Primer A: TGCTTTAAATTCCTCCCTGGA  
Primer B: AATGAGGCTCAGAGAAAGTGTTG  
STS size: 348  
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Amplifrag Gold Polymerase: 0.07 units/ul  
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES  
source  
1..496  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="10"  
/clone\_lib="Human"  
132..479  
132..154  
primer\_bind  
primer\_bind complement(457..479)  
BASE COUNT 129 a 103 c 128 g 133 t 3 others  
ORIGIN

Query Match 2.0%; Score 69; DB 11; Length 496;  
Best Local Similarity 100.0%; Pred. No. 9.2e-28;  
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 630 GGAATTCCTTGGATTCATTTTCAACCAAGGAGTGAAGCAAGATCAAGAGGAT 689  
DB 21 GGAATTCCTTGGATTCATTTTCAACCAAGGAGTGAAGCAAGATCAAGAGGAT 80  
QY 690 GGTTCCTCA 698  
DB 81 GGTTCCTCA 89

G50805 496 bp DNA linear STS 30-MAR-2000  
SHGC-83858 Human Homo sapiens STS genomic, sequence tagged site.  
G50805  
G50805.1 GI:5221982  
STS.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 496)  
Olivier, M. and Cox, D.R.  
Unpublished, Olivier, M., Cox, D.R. (2000)  
Unpublished (2000)

Contact: Michael Olivier, David R. Cox  
Stanford Human Genome Center  
Stanford University School of Medicine  
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA  
Tel: (650) 320-5800  
Fax: (650) 320-5801  
Email: olivier@shgc.stanford.edu  
Primer A: TGCTTTAAATTCCTCCCTGGA  
Primer B: AATGAGGCTCAGAGAAAGTGTTG  
STS size: 348  
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Amplifrag Gold Polymerase: 0.07 units/ul  
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES  
source  
1..496  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="10"  
/clone\_lib="Human"  
132..479  
132..154  
primer\_bind  
primer\_bind complement(457..479)  
BASE COUNT 129 a 103 c 128 g 133 t 3 others  
ORIGIN

Query Match 2.0%; Score 69; DB 11; Length 496;  
Best Local Similarity 100.0%; Pred. No. 9.2e-28;  
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 630 GGAATTCCTTGGATTCATTTTCAACCAAGGAGTGAAGCAAGATCAAGAGGAT 689  
DB 21 GGAATTCCTTGGATTCATTTTCAACCAAGGAGTGAAGCAAGATCAAGAGGAT 80  
QY 690 GGTTCCTCA 698  
DB 81 GGTTCCTCA 89

```

RESULT 5
AC125150      154748 bp      DNA      linear      HTG 26-JUN-2002
LOCUS         Mus musculus chromosome UNK clone RP24-354K9, WORKING DRAFT
DEFINITION    SEQUENCE, 8 unordered pieces.
ACCESSION     AC125150.1 GI:21490651
VERSION        HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS       house mouse.
SOURCE         Mus musculus
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1. (bases 1 to 154748)
AUTHORS        McPherson,J.D. and Waterston,R.H.
TITLE          The sequence of Mus musculus clone
JOURNAL         Unpublished
REFERENCE      2. (bases 1 to 154748)
AUTHORS        McPherson,J.D. and Waterston,R.H.
TITLE          Direct Submission
JOURNAL         Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
                Parkway, St. Louis, MO 63108, USA
REFERENCE      3. (bases 1 to 154748)
AUTHORS        McPherson,J.D. and Waterston,R.H.
TITLE          Direct Submission
JOURNAL         Submitted (26-JUN-2002) Genome Sequencing Center, 4444 Forest Park
                Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submission@wustl.wustl.edu
----- Project Information -----
Center project name: M_BB0354K09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151541 bases at least Q40
Consensus quality: 151970 bases at least Q30
Consensus quality: 152209 bases at least Q20
Insert size: 164000; agarose-fp
Insert coverage: 8.91 in Q20 bases;
Quality coverage: 8.71 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1201: contig of 1201 bp in length
* 1301: gap of unknown length
* 1302: contig of 2738 bp in length
* 4039: gap of unknown length
* 4040: contig of 31778 bp in length
* 35918: gap of unknown length
* 35918: contig of 4530 bp in length
* 36018: gap of unknown length
* 40547: gap of unknown length
* 40647: gap of unknown length
* 51635: contig of 10988 bp in length
* 51735: gap of unknown length
* 51636: contig of 35396 bp in length
* 51736: gap of unknown length
* 87131: contig of 66133 bp in length
* 87132: gap of unknown length
* 87232: contig of 1284 bp in length
* 153365: gap of unknown length
* 153465: contig of 1284 bp in length.

```

```

FEATURES
source
Location/Qualifiers
1. .154748
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-354K9"
misc_feature
1. .1201
/note="assembly_name:Contig10"
1302. .4039
/note="assembly_name:Contig11"
clone_end:SP6
vector_side:right
misc_feature
4140. .35917
/note="assembly_name:Contig14"
clone_end:T7
vector_side:right
misc_feature
36018. .40547
/note="assembly_name:Contig12"
40648. .51635
/note="assembly_name:Contig13"
51736. .87131
/note="assembly_name:Contig15"
87232. .153364
/note="assembly_name:Contig16"
153465. .154748
/note="assembly_name:Contig18"
BASE COUNT 39535 a 38195 c 37106 g 39199 t 713 others
ORIGIN

Query Match 0.9%; Score 31; DB 2; Length 154748;
Best Local Similarity 100.0%; Pred. No. 2.4e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 GTTCCACTCCTCATCTGGAATTCCTGGA 644
|||||
DB 95867 GTTCCACTCCTCATCTGGAATTCCTGGA 95897

RESULT 6
F243081S03      1728 bp      DNA      linear      PRI 13-JUN-2001
LOCUS         Homo sapiens intrinsic factor-vitamin B12 receptor (CUBN) gene,
DEFINITION    exons 5 and 6.
ACCESSION     AF243083
VERSION        AF243083.1 GI:14388625
KEYWORDS       3 of 49
SEGMENT       Homo sapiens.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1. (bases 1 to 1728)
AUTHORS        Aminoff,M., Brady,S., Verroust,P.J., Moestrup,S.K. and Krahe,R.
TITLE          The genomic structure of the human CUBN gene encoding cubilin, the
                intrinsic factor-vitamin B12 receptor
JOURNAL         Unpublished
REFERENCE      2. (bases 1 to 1728)
AUTHORS        Aminoff,M., Brady,S., Verroust,P.J., Moestrup,S.K. and Krahe,R.
TITLE          Direct Submission
JOURNAL         Submitted (08-MAR-2000) Division of Human Cancer Genetics, Ohio
                State University, 420 West 12th Avenue, Columbus, OH 43210, USA

FEATURES
source
Location/Qualifiers
1. .1728
/organism="Homo sapiens"
/db_xref="taxon:9606"
125. .226
/exon         /gene="CUBN"
/number=5
916. .1019
/exon         /gene="CUBN"
/number=6
498 a 351 c 347 g 532 t
BASE COUNT 498 a 351 c 347 g 532 t
ORIGIN

```



misc\_feature /note="encodes EGF repeat"  
666..815  
misc\_feature /note="encodes EGF repeat"  
816..944  
misc\_feature /note="encodes EGF repeat"  
945..1076  
misc\_feature /note="encodes EGF repeat"  
1077..1221  
misc\_feature /note="encodes EGF repeat"  
1222..1322  
misc\_feature /note="encodes EGF repeat"  
1323..1436  
misc\_feature /note="encodes EGF repeat"  
1446..1793  
misc\_feature /note="encodes CUB domain"  
1794..2147  
misc\_feature /note="encodes CUB domain"  
2148..2474  
misc\_feature /note="encodes CUB domain"  
2475..2819  
misc\_feature /note="encodes CUB domain"  
2820..3167  
misc\_feature /note="encodes CUB domain"  
3168..3518  
misc\_feature /note="encodes CUB domain"  
3519..3857  
misc\_feature /note="encodes CUB domain"  
3858..4196  
misc\_feature /note="encodes CUB domain"  
4197..4553  
misc\_feature /note="encodes CUB domain"  
4554..4883  
misc\_feature /note="encodes CUB domain"  
4884..5237  
misc\_feature /note="encodes CUB domain"  
5238..5582  
misc\_feature /note="encodes CUB domain"  
5583..5957  
misc\_feature /note="encodes CUB domain"  
5958..6299  
misc\_feature /note="encodes CUB domain"  
6300..6674  
misc\_feature /note="encodes CUB domain"  
6675..7031  
misc\_feature /note="encodes CUB domain"  
7032..7379  
misc\_feature /note="encodes CUB domain"  
7380..7733  
misc\_feature /note="encodes CUB domain"  
7734..8090  
misc\_feature /note="encodes CUB domain"  
8091..8438  
misc\_feature /note="encodes CUB domain"  
8439..8783  
misc\_feature /note="encodes CUB domain"  
8784..9134  
misc\_feature /note="encodes CUB domain"  
9135..9494  
misc\_feature /note="encodes CUB domain"  
9495..9857  
misc\_feature /note="encodes CUB domain"  
9858..10208  
misc\_feature /note="encodes CUB domain"  
10209..10556  
misc\_feature /note="encodes CUB domain"  
10557..10895

Query Match 0.7%; Score 25; DB 9; Length 11344;  
Best Local Similarity 100.0%; Pred. No. 0.069;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1236 CTGCCAGATGGAGGCACATGTGT 1260  
|||||

Db 575 CTGCCAGATGGAGGCACATGTGT 599  
RESULT 8  
AC067747  
LOCUS AC067747 147280 bp DNA linear PRI 24-APR-2002  
DEFINITION Homo sapiens chromosome 10 clone RP11-406H21, complete sequence.  
AC067747  
VERSION AC067747.7 GI:20279381  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 147280)  
AUTHORS Smith, D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome  
JOURNAL Sequence Data  
REFERENCE 2 (bases 1 to 147280)  
AUTHORS Smith, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
REFERENCE 3 (bases 1 to 147280)  
AUTHORS Smith, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2002) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
REFERENCE 4 (bases 1 to 147280)  
AUTHORS Smith, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (24-APR-2002) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
COMMENT On Apr 24, 2002 this sequence version replaced gi:18767391.  
FEATURES  
source  
location:Qualifiers  
1..147280  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-406H21"  
/clone\_lib="RPC1-11"  
BASE COUNT 43059 a 27509 c 27812 g 48900 t  
ORIGIN  
Query Match 0.7%; Score 25; DB 9; Length 147280;  
Best Local Similarity 100.0%; Pred. No. 0.078;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1236 CTGCCAGATGGAGGCACATGTGT 1260  
|||||  
Db 50565 CTGCCAGATGGAGGCACATGTGT 50589  
RESULT 9  
AC119647/c  
LOCUS AC119647 152895 bp DNA linear HTG 18-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-137H14, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 54 unordered pieces.  
AC119647  
VERSION AC119647.2 GI:21747172  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 152895)  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carton, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J., Earhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McWhiney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roife-Mc, Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 152895)  
Worley, K.C.  
Direct Submission  
Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 152895)  
Worley, K.C.  
Direct Submission  
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20340380.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWPZ  
Center clone name: CH230-137H14  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 91714 bases at least Q40  
Consensus quality: 98583 bases at least Q30  
Consensus quality: 103878 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 54 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1041: contig of 1041 bp in length  
1042 1141: gap of unknown length  
1142 2262: contig of 1121 bp in length  
2263 2362: gap of unknown length  
2363 3497: contig of 1135 bp in length  
3498 3597: gap of unknown length  
3598 5210: contig of 1613 bp in length  
5211 5310: gap of unknown length  
5311 6936: contig of 1626 bp in length  
6937 7037: contig of unknown length  
7037 8291: contig of 1255 bp in length  
8292 9663: contig of 1272 bp in length  
9664 9763: gap of unknown length  
9764 11369: contig of 1606 bp in length  
11370 11469: gap of unknown length  
11470 13073: contig of 1604 bp in length  
13074 13173: gap of unknown length  
13174 14521: contig of 1348 bp in length  
14522 14621: gap of unknown length  
14622 16274: contig of 1653 bp in length  
16275 16374: gap of unknown length  
16374 17405: contig of 1031 bp in length  
17406 17505: gap of unknown length  
17506 18547: contig of 1042 bp in length  
18548 18647: gap of unknown length  
18648 19808: contig of 1261 bp in length  
19809 20008: gap of unknown length  
20009 21460: contig of 1452 bp in length  
21461 21560: gap of unknown length  
21561 22822: contig of 1262 bp in length  
22823 22922: gap of unknown length  
22923 24437: contig of 1515 bp in length  
24438 24537: gap of unknown length  
24538 25832: contig of 1295 bp in length  
25833 25932: gap of unknown length  
25933 27104: contig of 1172 bp in length  
27105 27204: gap of unknown length  
27205 29152: contig of 1848 bp in length  
29153 30398: contig of 1246 bp in length  
30399 30498: gap of unknown length  
30499 32426: contig of 1928 bp in length  
32427 32526: gap of unknown length  
32527 34955: contig of 2429 bp in length  
34956 35055: gap of unknown length  
35056 36842: contig of 1587 bp in length  
36843 36742: gap of unknown length  
36743 37820: contig of 1078 bp in length  
37821 37920: gap of unknown length  
37921 39394: contig of 1474 bp in length  
39395 39494: gap of unknown length  
39495 40928: contig of 1434 bp in length  
40929 41028: gap of unknown length  
41029 43380: contig of 2352 bp in length  
43381 43480: gap of unknown length  
43481 44748: contig of 1268 bp in length  
44749 44848: gap of unknown length  
44849 45927: contig of 2079 bp in length  
45928 47027: gap of unknown length  
47028 49481: contig of 2454 bp in length  
49482 49581: gap of unknown length  
49582 51923: contig of 2342 bp in length  
51924 52023: gap of unknown length  
52024 54126: contig of 2103 bp in length  
54127 54226: gap of unknown length  
54227 57214: contig of 2988 bp in length  
57215 57314: gap of unknown length  
57315 60024: contig of 2710 bp in length  
60025 60124: gap of unknown length

\* 60125 63163: contig of 3039 bp in length  
 \* 63164 63263: gap of unknown length  
 \* 63264 66703: contig of 3440 bp in length  
 \* 66704 66803: gap of unknown length  
 \* 66804 68557: contig of 1754 bp in length  
 \* 68558 68657: gap of unknown length  
 \* 68658 72882: contig of 4225 bp in length  
 \* 72883 72982: gap of unknown length  
 \* 72983 76542: contig of 3560 bp in length  
 \* 76543 76642: gap of unknown length  
 \* 76643 80420: contig of 3778 bp in length  
 \* 80421 80520: gap of unknown length  
 \* 80521 84990: contig of 4470 bp in length  
 \* 84991 85090: gap of unknown length  
 \* 85091 88933: contig of 3843 bp in length  
 \* 88934 89033: gap of unknown length  
 \* 89034 96304: contig of 7271 bp in length  
 \* 96305 96404: gap of unknown length  
 \* 96405 101068: contig of 4664 bp in length  
 \* 101069 101168: gap of unknown length  
 \* 101169 105041: contig of 3873 bp in length  
 \* 105042 105141: gap of unknown length  
 \* 105142 110814: contig of 5673 bp in length  
 \* 110815 110914: gap of unknown length  
 \* 110915 115458: contig of 4544 bp in length  
 \* 115459 120911: contig of 5353 bp in length  
 \* 120912 121011: gap of unknown length  
 \* 121012 128307: contig of 7296 bp in length  
 \* 128308 128407: gap of unknown length  
 \* 128408 134219: contig of 5812 bp in length  
 \* 134220 134319: gap of unknown length  
 \* 134320 139473: contig of 5154 bp in length  
 \* 139474 139573: gap of unknown length  
 \* 139574 146170: contig of 6597 bp in length  
 \* 146171 146270: gap of unknown length  
 \* 146271 152895: contig of 6625 bp in length.

Query Match 0.7%; Score 25; DB 2; Length 152895;  
 Best Local Similarity 100.0%; Pred. No. 0.078;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 GTTCCACTCTCATCTCGAATTC 538  
 Db 122055 GTTCCACTCTCATCTCGAATTC 122031

RESULT 10  
 HSNOTCH2 HSNOTCH2 622 bp DNA linear PRI 05-APR-1995  
 LOCUS H.sapiens NOTCH 2 gene.  
 DEFINITION X80115  
 ACCESSION X80115.1 GI:763114  
 VERSION notch2 gene.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 622)  
 AUTHORS Lardelli, M.T.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JUL-1994) M.T. Lardelli, Karolinska Institute, Developmental Biology Laboratory, CMB, 171 77 Stockholm, SWEDEN  
 REFERENCE 2 (bases 1 to 622)  
 AUTHORS Larsson, C., Lardelli, M., White, I. and Lendahl, U.  
 TITLE The human NOTCH1, 2, and 3 genes are located at chromosome positions 9q34, 1p13-p11, and 19p13.2-p13.1 in regions of neoplasia-associated translocation  
 JOURNAL Genomics 24 (2), 253-258 (1994)  
 MEDLINE 95213014  
 PUBMED 7698746  
 FEATURES Location/Qualifiers  
 source 1..622

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="p13-p11"  
 /clone="B-11"  
 /sub\_clone="mc12"  
 /haplotype="diploid"  
 /tissue\_type="placental"  
 /clone\_lib="human placental cosmid library in pWE15 vector"  
 185..317  
 /gene="NOTCH2"  
 185..317  
 /gene="NOTCH2"  
 /note="EGF repeat 31"  
 167 a 122 c 132 g 201 t  
 BASE COUNT  
 ORIGIN  
 Query Match 0.7%; Score 23; DB 9; Length 622;  
 Best Local Similarity 100.0%; Pred. No. 0.89;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1231 CAGCCCTGCCAGATGAGGCAC 1253  
 Db 242 CAGCCCTGCCAGATGAGGCAC 264  
 RESULT 11  
 HSNOTCH2 HSNOTCH2 789 bp mRNA linear PRI 21-NOV-1996  
 LOCUS Human Notch2 mRNA, partial cds.  
 DEFINITION U77493  
 ACCESSION U77493.1 GI:1679773  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 789)  
 AUTHORS Lemasson, I., Devaux, C. and Mesnard, J.M.  
 TITLE Partial sequence of EGF-like repeat domain of human Notch2 mRNA  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 789)  
 AUTHORS Lemasson, I., Devaux, C. and Mesnard, J.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-NOV-1996) Laboratoire d'Immunologie des Infections Retrovirales, Institut de Biologie, 4, Boulevard Henri IV, Montpellier 34060 Cedex, France  
 FEATURES Location/Qualifiers  
 source 1..789  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="MT2"  
 /tissue\_type="T cell"  
 <1..>789  
 /gene="Notch2"  
 <1..>789  
 /gene="Notch2"  
 /note="transmembrane protein; hn"  
 /codon\_start=1  
 /product="Notch2"  
 /protein\_id="AAB19224.1"  
 /db\_xref="GI:1679774"  
 /translation="YVNSYTKCGAGFDGVCHEENINTESSCFNGGTCVDGINSFS  
 CLCPVGTGTFCLHEINECSHPCLNDGTCVDGLGYRCSCPLGYTGKNCOTLVNLC  
 RSPCKNGKTCVKKSCQICPSGWWGACDVPNVCDAASRGVILVHLCQHSVC  
 INAGNTHYCCQVPGTYGTCYCEQLDECASNPCHGATCSDFITGGYRCECVPGYGVNC  
 EYEVEDECNQPCQNGTCTDLVNHFKCSCPPGTRGLLCEINIDDC"  
 180 a 192 c 214 g 203 t  
 BASE COUNT  
 ORIGIN  
 Query Match 0.7%; Score 23; DB 9; Length 789;  
 Best Local Similarity 100.0%; Pred. No. 0.9;

```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 CAGCCCTGCAGAAATGAGGCAC 1253
|||||
Db 682 CAGCCCTGCAGAAATGAGGCAC 704

RESULT 12
HSA296168          924 bp  mRNA  linear  PRI 31-JAN-2001
LOCUS              Homo sapiens mRNA for keratin associated protein (KRTAP4.13 gene).
DEFINITION
ACCESSION          AJ296168
VERSION            1 GI:12655381
KEYWORDS            keratin associated protein; KRTAP4.13 gene.
SOURCE              human.
ORGANISM            Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Rogers M.A., Langbein, L., Winter, H., Ehmann, C., Korn, B. and
                    Schweizer, J.
TITLE              Characterization of a cluster of human high/ ultrahigh keratin
                    associated proteins on chromosome 17q12-21
JOURNAL
AUTHORS            Rogers M.A.
TITLE              Direct Submission
JOURNAL            Submitted (16-OCT-2000) Rogers M.A., Research Program B, German
                    Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg,
                    Germany 69120, GERMANY
COMMENT            Related genomic sequence: AC006070 (146419-146919nt).
FEATURES
                    Location/Qualifiers
                    1..1099
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="17"
                        /map="17q12-21"
                        /clone="sc123"
                        /tissue_type="scalp"
                        /gene="KRTAP4.4"
                        /protein_id="CAC27575.1"
                        /translation="MVNSCCGVCSDQGGLENCRPSYQCTTCRTTCRPSCCVSS
                        CCRPQCQTTCRTTCCHPSCCVSCCRPQCQSVCCQPTCCRPQCQTTCRTTCRPSCCVSRCPHC
                        GOSLCC"
                    polyA_signal 1058..1063
                    BASE COUNT 236 a 365 c 227 g 271 t
                    ORIGIN
Query Match          0.7%; Score 23; DB 9; Length 1099;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 CCACCTGCTACAGGACCACTGC 1212
|||||
Db 558 CCACCTGCTACAGGACCACTGC 580

RESULT 14
AF308601          9722 bp  mRNA  linear  PRI 22-NOV-2000
LOCUS              Homo sapiens NOTCH 2 (N2) mRNA, complete cds.
DEFINITION
ACCESSION          AF308601 M99437
VERSION            AF308601.1 GI:11275977
KEYWORDS
SOURCE              Homo sapiens.
ORGANISM            Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Stifani, S., Blaumueller, C.M., Redhead, N.J., Hill, R.E. and
                    Artavanis-Tsakonas, S.
TITLE              Human homologs of a Drosophila Enhancer of split gene product
                    define a novel family of nuclear proteins
JOURNAL            Nat. Genet. 2 (2), 119-127 (1992)
MEDLINE            93265135
PUBMED            1303260
REFERENCE
AUTHORS            Blaumueller, C.M. and Mann, R.S.
TITLE              Complete Human Notch 2 (hN2) cDNA sequence
JOURNAL            Unpublished
AUTHORS            Blaumueller, C.M. and Artavanis-Tsakonas, S.
REFERENCE
AUTHORS            Blaumueller, C.M. and Artavanis-Tsakonas, S.

```

TITLE Direct Submission  
JOURNAL Submitted (27-APR-1993) Boyer Center for Molecular Medicine, Howard Hughes Medical Institute, 295 Congress Ave., New Haven, CT 06519, USA

REFERENCE 4 (bases 1 to 9722)  
AUTHORS Blaumüller, C.M. and Mann, R.S.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-2000) Developmental Biology and Cancer, Massachusetts General Hospital Cancer Center, Building 149, 13th Street, Charlestown, MA 02129-2000, USA

REMARK Sequence update by submitter  
COMMENT On Nov 22, 2000 this sequence version replaced gi:189263.  
FEATURES Location/Qualifiers  
1..9722  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="female"  
/tissue\_type="brain"  
/dev\_stage="17-18 week embryo"  
1..9722  
/gene="N2"  
13..7428  
/gene="N2"  
/codon\_start=1  
/product="NOTCH 2"  
/protein\_id="AAA36377.2"  
/db\_xref="GI:1175978"  
/translation="MPALRPALLWALLLALCCAAHAALQCRDGYEPCVNEGMCVTY HNGYCKPGEFLGECYCOHRDCEKNCQNGCTCAVAMLGKATCASCAGFTGEDCQ YTSHPCFVSRPCLNGGTCMLSDRTYECQVGTGKEQWTDACLSPHCANGSTCT TVANQFCKCLGTGQKQEDVNECDIPGHQGGTCLNLPGSYQCOCQPGFTGYC DSYVPCAPFCVNGTCRQTDGTFECNCLPGFEGSTCERNIDDCPHRCQNGGVC DGVNTYCRCPQWGTQCTEDVDECLLPNAGCGGTCAHRNGGYGCVCVNGSGDD CSENIDCAFSCPTPGSTCDIRVASFCMCPPEGKAGLLCHLDACIINPCHGKALCPT NPLNGOYICTCPOGYKADCTEDVDECAMANSNPECHAGKCVNTDGAFCCEKLGKAG PCRMENRCHSDPCQNDATCLDKIGGFTCLMPGFKGVCELEINEQCNSPCVNNQGV CVDKVNRECLCPGPGFTGPVQCQDIDDCSTPCLNGAKCIDHPNGYECQCATGFTGVL CENINDCPDPCPHGQCGQDIDSYTCINPGYMGATCSDIDEICSSPCNLNDGRCID LVNGYQCQPGTSGVANCEINFDCAFNPCITHGICMDGINRYSCVSPGFTGRCNID IDEACSNPKRGATCINGVNGFCICPEGPHHPCSYSOVNECLSNPCINHGCTGLSG YKCLDAGWGVINEVDNNECLSNPCQNGGTCNLVNGYRCKCKGKGYNCOVNIDE CASNPLNGCTCFDIDSGYTCVCHVLPYTGNCQTVLAPCSFNPENAAVCKESNPFES YTCCLAPWGQRCQDIDIDEICSKPQNMHGLCHNTQGSYMCCECPGFGSMDCEDIDD CLANPCQNGSCMDVNTFSCILPGLFTGDKQTDNNECLSEPCNKGSTGSDYVNST CKCQAGFDGVHCENINIECTESSCFNGTCDVGINSPCLCPVGTGFTGFCGLHEINECS SHPCLNBTCTVDGLTVRCSCPLGTCNKQTLVNLCSRSCKNKGTCVOKKAEQSLC CPSCWAGACDVPWNSCDIAASRGVLVHLCOHSGVCINAGNTHYCCQPLGYTGVC EQULDEACSNPCQHGATCDFIGYRCYCPGVQGVNCEYEDVCEQPCQNGGTCID LVNHFKSCSPGTRGLICEENIDDCARGPHCLNGGQCMRIGGYSRCLPFGFAGERCE GDINECLSNPCSSSEGLDCIQLTNDYLCVCSAFTGRHCEYFDVCPQMPCLNGGTCA VASNMPDGFICRCPGFGARQSCQGVKCRKEQCVHTASGPRCFCPSRCESECC ASSPCQHGGSCHPOROPPYSCCAPPFGSRCELYTAPPSTPATCLSOYCADKARD GVDCACNSHACQWDGDCSLTWMENPWANCSSPLPCWDYINNOCDCLNTVECLDFNF ECOGNSKTCYDKYCADHFDNHCNQCNEECGWDGLDCAADQENLARGTLVIVVL MPPEQLQDARSELRALGTLNHLNLRKRSQGLMVPYIYGKSAAMKORMTRRSL PGBQEQVAVGSVFLETDNRQCVQSDHCFKNTDAAAALLASHAIOGLTSLPYLVSVS ESTPRTQTLVLAVAVVIILFILLIGVIMAKRKHGKGLWLPGLFTLRDASNNHR REPVGQDANGLNLSGVSEANLIGTCTSEHWVDEGPQPKVKADEALLSEDDPT DRPWTQOHLLEADIRETPSIALTPQAEQVDVLDVNVGPDGCTPLMLASLRGSS DLSEDEDESSNITDLYQASLOAQTDRTGEMALHLAARYSRADAARLLLDAG ADANAQDNMGRCPLHAAVAADAGVFOILIRNVRTDLDARNDGTTPLILAAALVAG MYAELINCOADVADVDDHGSALHWAANNVNEATLLLLKNGANRMDQNKETPLFL AAREGYSYAAKILLDHFANRDITDMDRLPRDVARDRMHDIVRLDLEYNVTSPGPT VLSALSPTVICPNRFLSLKHTPMGKSRPSAKSTMTSLPDLAKDAKGSRRK KLSLEKVLSESVTLSPVDSLESPTYVSDTTSSPMITSPGILQASPNMLATAAPP APVHAQHALSESNLHENOPLAHGASTVLPVSVQLLSHHHIVSGSGSAGLSRLHPVP VPADMMRMEVNETQYNEMFGWVLAPEGTHPGIAPOSRPPRKHITTPREPLPIVT FOLIPKGSIAQPCAGAPQCTPPAVAGPLPTWQIPEMARLPSVAFTAMFPQDQGV VAOTILPAYHPFPASVGKYPPTPSOHSYASSNAARTPSHSHLQGEHPHYLTSPESP DOWSSSSPHSADSDVTTTPTPGAGGGQGRGPGCTHMSPPHNNMQVYA"  
13..81  
/gene="N2"  
82..4257  
sig\_peptide  
misc\_feature

misc\_feature /gene="N2"  
/note="Region: EGF repeats"  
4291..4623  
/gene="N2"  
/note="Region: LIN12 repeats"  
5050..5124  
/gene="N2"  
/note="Region: transmembrane domain"  
5449..6525  
/gene="N2"  
/note="Region: CDC10 domain"  
5482..6123  
/gene="N2"  
/note="Region: CDC10/ANK repeats"  
7171..7356  
/gene="N2"  
/note="Region: PEST domain"  
/note="Region: 2451 g 2491 t  
BASE COUNT 2262 a 2518 c 2451 g 2491 t  
ORIGIN

Query Match 0.7%; Score 23; DB 9; Length 9722;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1231 CAGCCCTGCCAGATGGAGGCAC 1253  
|||||  
Db 3592 CAGCCCTGCCAGATGGAGGCAC 3614  
|||||

RESULT 15  
ARI02329  
LOCUS Sequence 21 from patent US 6083904. linear PAT 14-FEB-2001  
DEFINITION ARI02329  
ACCESSION ARI02329  
VERSION ARI02329.1 GI:12813127  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 9723)  
AUTHORS Artavanis-Tsakonas, S.  
TITLE Therapeutic and diagnostic methods and compositions based on notch proteins and nucleic acids  
JOURNAL Patent: US 6083904-A 21 04-JUL-2000;  
FEATURES Location/Qualifiers  
1..9723  
source /organism="unknown"  
BASE COUNT 2262 a 2520 c 2452 g 2489 t  
ORIGIN

Query Match 0.7%; Score 23; DB 6; Length 9723;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1231 CAGCCCTGCCAGATGGAGGCAC 1253  
|||||  
Db 3586 CAGCCCTGCCAGATGGAGGCAC 3608  
|||||

Search completed: May 5, 2003, 07:30:22  
Job time : 8640 secs

**THIS PAGE BLANK (USPTO)**

Result No.	Query Match	Score	Length	DB	ID	Description
1	100.0	3375	24	ABK92207	Prostate cancer-as	
2	1537	45.5	3485	23	AA815130	DNA encoding novel
3	296	8.8	406	23	AA815130	DNA encoding novel
4	32	0.9	1837	21	AA261803	CDNA encoding muri
5	32	0.9	1837	22	AAC99736	Skin cell cDNA, SE
6	32	0.9	1837	24	ABL34888	Murine cDNA isolat
7	25	0.7	1422	23	AA92463	DNA encoding novel
8	25	0.7	11344	23	AA92463	DNA encoding novel
9	32	0.7	394	21	AAC05374	Human secreted pro

PR 04 - MAY - 2001; 2001US-288589P.  
XX



```
QY 1621 CGTGGCTTCGGAGCGCCACAGCAGCAGCCAGGACCGCCGACGCTAGAGTGGTGGTTTG 1680
Db 1621 CGTGGCTTCGGAGCGCCACAGCAGCAGCCAGGACCGCCGACGCTAGAGTGGTGGTTTG 1680
QY 1681 CTCAGTGAATCACACTCCGAGGATGAGTTGGGGCCCGAGCGCTACGCAAGGCGCGA 1740
Db 1681 CTCAGTGAATCACACTCCGAGGATGAGTTGGGGCCCGAGCGCTACGCAAGGCGCGA 1740
QY 1741 GAGTGTCTCTGCTGGGTGTAGGAGTGAAGCCGCTGCGGGCAGAGCTGGAGGAGATCACA 1800
Db 1741 GAGTGTCTCTGCTGGGTGTAGGAGTGAAGCCGCTGCGGGCAGAGCTGGAGGAGATCACA 1800
QY 1801 GGCAGCCCAAGATGATGGTCTACTCGGATCCTCAGGATCTGTTCACCAAAATCCCT 1860
Db 1801 GGCAGCCCAAGATGATGGTCTACTCGGATCCTCAGGATCTGTTCACCAAAATCCCT 1860
QY 1861 GAGCTGCAGGGAAGCTGTGAGCGCGGCGAGCGCCGAGGCTGCCGACACAGCCCTGGAC 1920
Db 1861 GAGCTGCAGGGAAGCTGTGAGCGCGGCGAGCGCCGAGGCTGCCGACACAGCCCTGGAC 1920
QY 1921 CTCGTCTTCATGTTGGACACCTCTGCCCTCAGTAGGGCCCGAGAAATTTTGTCTCAGATGCG 1980
Db 1921 CTCGTCTTCATGTTGGACACCTCTGCCCTCAGTAGGGCCCGAGAAATTTTGTCTCAGATGCG 1980
QY 1981 AGCTTTGTGAGAAGCTGTGCCCTCAGTTTGAAGTGAACCTGACGTGACACAGTCCGGC 2040
Db 1981 AGCTTTGTGAGAAGCTGTGCCCTCAGTTTGAAGTGAACCTGACGTGACACAGTCCGGC 2040
QY 2041 CTGGTGTGTATGACAGCCAGCTGACAGCTGCTTCGGGCTGGACACCAACCCACCCGG 2100
Db 2041 CTGGTGTGTATGACAGCCAGCTGACAGCTGCTTCGGGCTGGACACCAACCCACCCGG 2100
QY 2101 GCTGGATGCTCGCGGCCATTAGCCAGGCCCTACCTAGTGGGGTGGGCTCAGCCCGC 2160
Db 2101 GCTGGATGCTCGCGGCCATTAGCCAGGCCCTACCTAGTGGGGTGGGCTCAGCCCGC 2160
QY 2161 ACCGCCCTGTGCATCTATGACAAAGTATGACCGTCCAGAGGGTGGCCCGGCTGGT 2220
Db 2161 ACCGCCCTGTGCATCTATGACAAAGTATGACCGTCCAGAGGGTGGCCCGGCTGGT 2220
QY 2221 GTCCCCAAGCTGTGGTGGTCTACAGCGGAGGAGCGCAGAGGATGACAGCGCTTCC 2280
Db 2221 GTCCCCAAGCTGTGGTGGTCTACAGCGGAGGAGCGCAGAGGATGACAGCGCTTCC 2280
QY 2281 GCCCAGAGCTGAGSAACAATGGCATCTCTCTTGTGCTGGCGTGGGCGCTGTCCTA 2340
Db 2281 GCCCAGAGCTGAGSAACAATGGCATCTCTCTTGTGCTGGCGTGGGCGCTGTCCTA 2340
QY 2341 AGTGAGGCTGTGGAGGCTTGCAGGTCCTCCCGGATCCCTGTATCCAGTGGCAGCTTAC 2400
Db 2341 AGTGAGGCTGTGGAGGCTTGCAGGTCCTCCCGGATCCCTGTATCCAGTGGCAGCTTAC 2400
QY 2401 GCCGACCTGCGGTACACACGAGCTGCTCATTTGATGGCTGTGGAGAGCCCAAGCAG 2460
Db 2401 GCCGACCTGCGGTACACACGAGCTGCTCATTTGATGGCTGTGGAGAGCCCAAGCAG 2460
QY 2461 CCAGTCAACCTCTGCAAAACCCAGCCGCTGATGAATGAGGCGAGCTGGCTCTGCAGAA 2520
Db 2461 CCAGTCAACCTCTGCAAAACCCAGCCGCTGATGAATGAGGCGAGCTGGCTCTGCAGAA 2520
QY 2521 GGGAGCTACCCGCTGAAGTGTGGGATGCTGTGGAGGGCCCCCTGCGGAGAACCGTGAG 2580
Db 2521 GGGAGCTACCCGCTGAAGTGTGGGATGCTGTGGAGGGCCCCCTGCGGAGAACCGTGAG 2580
QY 2581 TGGAGCTCTTCTCTGTATGTGTAGCCAGGATGATTTGTGAGCCGCCCTGAGGCAC 2640
Db 2581 TGGAGCTCTTCTCTGTATGTGTAGCCAGGATGATTTGTGAGCCGCCCTGAGGCAC 2640
QY 2641 ATGGCTCCCGTGCAGGAGGGCAGAGCGTACCCCTCCAGCAACTACAGAGAAGGCGCTG 2700
Db 2641 ATGGCTCCCGTGCAGGAGGGCAGAGCGTACCCCTCCAGCAACTACAGAGAAGGCGCTG 2700
QY 2701 GGCACCTGAAATGGTGCCTTCTTGGAAATGCTGTGCCCCAGGCTCCTTAGAATGTCTGC 2760
```

```
Db 2701 GGCACCTGAAATGGTGCCTTCTTGGAAATGCTGTGCCCCAGGCTCCTTAGAATGTCTGC 2760
QY 2761 TTCCCGCGGTGGCCAGGACCACTATTCTCACTGAGGGAGGAGGATGCCCAACTGCAGCC 2820
Db 2761 TTCCCGCGGTGGCCAGGACCACTATTCTCACTGAGGGAGGAGGATGCCCAACTGCAGCC 2820
QY 2821 ATGCTGCTTAGAGACAAGAAAGCAGCTGATGTCACCCCAACAGATGTTGTTGAAAAGTT 2880
Db 2821 ATGCTGCTTAGAGACAAGAAAGCAGCTGATGTCACCCCAACAGATGTTGTTGAAAAGTT 2880
QY 2881 TTGATGTGTAGTAAATACCCACTTTCTGTACCTGCTGCTGCTGCTGCTGCTGCTGCTAT 2940
Db 2881 TTGATGTGTAGTAAATACCCACTTTCTGTACCTGCTGCTGCTGCTGCTGCTGCTGCTAT 2940
QY 2941 CTGCCACTTTCCCTTGAGGATAAACAAGGGTCTCTGAAGACTTAAATTTAGCGGCCCTGA 3000
Db 2941 CTGCCACTTTCCCTTGAGGATAAACAAGGGTCTCTGAAGACTTAAATTTAGCGGCCCTGA 3000
QY 3001 CGTTCTTTGCACAAATCAATGCTGCCAGAAATGTTGTTGACACAGTAATGCCAGAG 3060
Db 3001 CGTTCTTTGCACAAATCAATGCTGCCAGAAATGTTGTTGACACAGTAATGCCAGAG 3060
QY 3061 AGGCTTTTACTAGACATCCTTTTGGACGGCGAAGGCCACGGCTTCAAGATGGAAGCA 3120
Db 3061 AGGCTTTTACTAGACATCCTTTTGGACGGCGAAGGCCACGGCTTCAAGATGGAAGCA 3120
QY 3121 GCAGCTTTTCCACTTCCCAGAGACATTCGATGATGCATTTTCATTTGCTGAAAGGGG 3180
Db 3121 GCAGCTTTTCCACTTCCCAGAGACATTCGATGATGCATTTTCATTTGCTGAAAGGGG 3180
QY 3181 CTTGAGGAGACCTTTTGTGACTTCTTTGGGACCTGCTTTTGTGTGGAGAGACTTGGAAA 3240
Db 3181 CTTGAGGAGACCTTTTGTGACTTCTTTGGGACCTGCTTTTGTGTGGAGAGACTTGGAAA 3240
QY 3241 GGTCTCAGACTGAATGTGACCAATTAACCACTTGGTTGATGATGGGGAGGGCTGAGT 3300
Db 3241 GGTCTCAGACTGAATGTGACCAATTAACCACTTGGTTGATGATGGGGAGGGCTGAGT 3300
QY 3301 TGTGCATGGGGCCAGGCTCTGAGGGCCACGTAAATCCCTTCTGAGTCTGTGAGCAGTGCC 3360
Db 3301 TGTGCATGGGGCCAGGCTCTGAGGGCCACGTAAATCCCTTCTGAGTCTGTGAGCAGTGCC 3360
QY 3361 ACCTTGAAGGCTTTC 3375
Db 3361 ACCTTGAAGGCTTTC 3375
```

## RESULT 2

```
AAS81530
ID AAS81530 standard; cDNA; 3485 BP.
XX AC AAS81530;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #17334.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN W0200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
```

```
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG17343.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 1; SEQ ID No 17334; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequencing data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 3485 BP; 829 A; 850 C; 1030 G; 776 T; 0 other;

Query Match 45.5%; Score 1537; DB 23; Length 3485;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1215 AGGCCCTGTGACTCGCAGCCCTGCCAGAAATGGAGGCACATGTGTCCAGAAGACTGGA 1274
|
|
|
DB 1949 AGGCCCTGTGACTCGCAGCCCTGCCAGAAATGGAGGCACATGTGTCCAGAAGACTGGA 2008
|
|
|
QY 1275 CGGTACCACTGCTCTGCCCGCTGGCCTTTGGAGGGAGGCTTAACGTGCCCTGAAGCT 1334
|
|
|
DB 2009 CGGCTACCACTGCTCTGCCCGCTGGCCTTTGGAGGGAGGCTTAACGTGCCCTGAAGCT 2068
|
|
|
QY 1335 GAGCCTGGAATGAGGGTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1394
|
|
|
DB 2069 GAGCCTGGAATGAGGGTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2128
|
|
|
QY 1395 GGACGGCTTCTCGGGGCCAAAGTCTTCGTGAAGCGGTTTGTGGGCGCTGCTGAGCGA 1454
|
|
|
DB 2129 GGACGGCTTCTCGGGGCCAAAGTCTTCGTGAAGCGGTTTGTGGGCGCTGCTGAGCGA 2188
|
|
|
QY 1455 GGACTCTCGGGCCGAGTGGGTGTGGCCACATACAGCAGGAGCTGCTGGTGGCGGTGCC 1514
|
|
|
DB 2189 GGACTCTCGGGCCGAGTGGGTGTGGCCACATACAGCAGGAGCTGCTGGTGGCGGTGCC 2248
|
|
|
QY 1515 TGTGGGGAGTACAGGATGTGCTGACCTGGTCTGGAGCCTGATGGCATTCCTTCGG 1574
|
|
|
DB 2249 TGTGGGGAGTACAGGATGTGCTGACCTGGTCTGGAGCCTGATGGCATTCCTTCGG 2308
|
|
|
QY 1575 TGTGGGCCCCACCTGACGGGCAGTGCCTTGGCGCAGGGCGGACAGCGTGGCTTCGGGAG 1634
|
|
|
DB 2309 TGTGGGCCCCACCTGACGGGCAGTGCCTTGGCGCAGGGCGGACAGCGTGGCTTCGGGAG 2368
|
|
|
QY 1635 CGCCACAGGACAGGCCAGGACCGGCCACGTAGAGTGGTGGTTTTGCTCACTAGATCACA 1694
|
|
|
DB 2369 CGCCACAGGACAGGCCAGGACCGGCCACGTAGAGTGGTGGTTTTGCTCACTAGATCACA 2428
|
|
|

1695 CTCGAGGATGAGGTTGCGGGGCCAGCGGCTCAGCAAGGGCGGAGAGCTGCTCTGCT 1754
|
|
|
2429 CTCGAGGATGAGGTTGCGGGGCCAGCGGCTCAGCAAGGGCGGAGAGCTGCTCTGCT 2488
|
|
|
1755 GGTGTAGGAGTGGAGCGCTGCGGGCAGAGCTGGAGAGATCACAGGACAGCCCAAGCA 1814
|
|
|
2489 GGTGTAGGAGTGGAGCGCTGCGGGCAGAGCTGGAGAGATCACAGGACAGCCCAAGCA 2548
|
|
|
1815 TGTGATGGTCTACTCGGATCCTCAGGATCTGTTCAACAAATCCCTGAGCTGAGGGGAA 1874
|
|
|
2549 TGTGATGGTCTACTCGGATCCTCAGGATCTGTTCAACAAATCCCTGAGCTGAGGGGAA 2608
|
|
|
1875 GCTGTGAGCGCGCAGCGGCGGAGGTCGCGGACACAAAGCCCTGAGCTCTCTCATGTT 1934
|
|
|
2609 GCTGTGAGCGCGCAGCGGCGGAGGTCGCGGACACAAAGCCCTGAGCTCTCTCATGTT 2668
|
|
|
1935 GGACACTCTGCGCTCAGTAGGCGCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAG 1994
|
|
|
2669 GGACACTCTGCGCTCAGTAGGCGCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAG 2728
|
|
|
1995 CTGTGCGCTCCAGTTTGAAGTGAACCTGACGTGACACAGCTGGCGCTGGTGTATGG 2054
|
|
|
2729 CTGTGCGCTCCAGTTTGAAGTGAACCTGACGTGACACAGCTGGCGCTGGTGTATGG 2788
|
|
|
2055 CAGCCAGGTGCACTGCGCTTCGGGTGGACACCAACCCACCGGGCTCGGATGCTGG 2114
|
|
|
2789 CAGCCAGGTGCACTGCGCTTCGGGTGGACACCAACCCACCGGGCTCGGATGCTGG 2848
|
|
|
2115 GGCATTAGCCAGCCCGCTTACCTAGTGGGTGGGTGAGCGGCGGACCCGCTGCTGCA 2174
|
|
|
2849 GGCATTAGCCAGCCCGCTTACCTAGTGGGTGGGTGAGCGGCGGACCCGCTGCTGCA 2308
|
|
|
2175 CATCTATGACAAAGTGATGACCGTCCAGAGGGGTGCCCGGCTGGTGTCCCCAAAGCTGT 2234
|
|
|
2909 CATCTATGACAAAGTGATGACCGTCCAGAGGGGTGCCCGGCTGGTGTCCCCAAAGCTGT 2968
|
|
|
2235 GGTGTGCTCAGAGCGGGAGAGCGGCGGAGGATGAGCGGCTTCTGCGCCAGAGCTGAG 2294
|
|
|
2969 GGTGTGCTCAGAGCGGGAGAGCGGCGGAGGATGAGCGGCTTCTGCGCCAGAGCTGAG 3028
|
|
|
2295 GAACAAATGGCATCTCTGTCTTGGTCTGCGGTGGGCGCTGTCTAAGTCAGGGTCTGCG 2354
|
|
|
3029 GAACAAATGGCATCTCTGTCTTGGTCTGCGGTGGGCGCTGTCTAAGTCAGGGTCTGCG 3088
|
|
|
2355 GAGGCTTGAGGTCCTCGGGATTCCTGTATCCAGTGGGAGCTTACGCCACCTCGCGGTA 2414
|
|
|
3089 GAGGCTTGAGGTCCTCGGGATTCCTGTATCCAGTGGGAGCTTACGCCACCTCGCGGTA 3148
|
|
|
2415 CCACCAAGGAGTCTCATTTAGTGGGTGTGTGGAGAGCCCAAGCCAGCTCAACCTCTG 2474
|
|
|
3149 CCACCAAGGAGTCTCATTTAGTGGGTGTGTGGAGAGCCCAAGCCAGCTCAACCTCTG 3208
|
|
|
2475 CAACCAAGGAGTCTCATTTAGTGGGTGTGTGGAGAGCCCAAGCCAGCTCAACCTCTG 2534
|
|
|
3209 CAACCAAGGAGTCTCATTTAGTGGGTGTGTGGAGAGCCCAAGCCAGCTCAACCTCTG 3268
|
|
|
2535 CAAGTGTGGGATGGGTGGAGGGCCCCACTCGGAGAACCGGTGAGTGGAGCTTTGCTC 2594
|
|
|
3269 CAAGTGTGGGATGGGTGGAGGGCCCCACTCGGAGAACCGGTGAGTGGAGCTTTGCTC 3328
|
|
|
2595 TGTATGTGAGCCAGGATGATTTGAGAGCGCCCTGAGGCACATGCTCCCGTGA 2654
|
|
|
3329 TGTATGTGAGCCAGGATGATTTGAGAGCGCCCTGAGGCACATGCTCCCGTGA 3388
|
|
|
2655 GGAGGCGAGCAGCCGCTACCCCTCCCAACAACTACAGAGAGGCTTGGGCACTGAAATGGT 2714
|
|
|
3389 GGAGGCGAGCAGCCGCTACCCCTCCCAACAACTACAGAGAGGCTTGGGCACTGAAATGGT 3448
|
|
|
2715 GCCTACTTCTGGAATGTGTGTGCCCCAGGTCTTAG 2751
|
|
|
3449 GCCTACTTCTGGAATGTGTGTGCCCCAGGTCTTAG 3485
|
|
|
```

## RESULT 3

AA581529/c  
ID AAS81529 standard; cDNA; 406 BP.  
XX  
AC AAS81529;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17333.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB; ABG17342.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID NO 17333; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependant on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 406 BP; 80 A; 112 C; 107 G; 107 T; 0 other;

Query Match 8.8%; Score 296; DB 23; Length 406;  
Best Local Similarity 100.0%; Pred. No. 2.1e-131;  
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2616 GATTCTTTGAGACGCCCTGAGGCACATGGCTCCCGTGCAGGAGGGCAGCGTACCC 2675  
|||||  
DB 312 GATTCTTTGAGACGCCCTGAGGCACATGGCTCCCGTGCAGGAGGGCAGCGTACCC 253  
|||||

QY 2676 TCCAGCAACTACAGAGAGCCCTGGCACTGAATGGTGCCTTCTTGGAAATGCTG 2735  
|||||  
DB 252 TCCAGCAACTACAGAGAGCCCTGGCACTGAATGGTGCCTTCTTGGAAATGCTG 193  
|||||

QY 2736 TGCCCCAGGTCCTTAGAATCTCTGCTTCCCGCGTGCCAGGACCACCTATTCTCAGTGAG 2795  
|||||  
DB 192 TGCCCCAGGTCCTTAGAATCTCTGCTTCCCGCGTGCCAGGACCACCTATTCTCAGTGAG 133  
|||||  
QY 2796 GGAGGAGGATGCCCACTGCAGCCATGCTTAGAGACAAGAAACACGCTGATGTCAC 2855  
|||||  
DB 132 GGAGGAGGATGCCCACTGCAGCCATGCTTAGAGACAAGAAACACGCTGATGTCAC 73  
|||||  
QY 2856 CCACAAACGATGTTGTTGAAAGTTTTGATGTGTAGTAATAATACCCACTTCTGTA 2911  
|||||  
DB 72 CCACAAACGATGTTGTTGAAAGTTTTGATGTGTAGTAATAATACCCACTTCTGTA 17  
|||||

## RESULT 4

AAZ61803  
ID AAZ61803 standard; cDNA; 1837 BP.  
XX  
AC AAZ61803;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE cDNA encoding murine skin cell secreted protein, SEQ ID NO:350.  
XX  
KW Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;  
KW embryonic skin cell; keratinocyte stem cell; transit amplifying cell;  
KW secreted; transmembrane; inflammation; cancer; neurological disease;  
KW angiogenesis; tumour vascularisation; growth disorder;  
KW developmental disorder; skin wound; hair follicle disorder;  
KW anti-inflammatory; cytostatic; neuroprotective; vulnery; ss.  
XX  
OS Mus sp.  
XX  
PN W09955865-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 29-APR-1999; 99WO-NZ00051.  
XX  
PR 29-APR-1998; 98US-0069726.  
PR 09-NOV-1998; 98US-0188930.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
PI Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murlison JG;  
XX WPI: 2000-072177/06.  
XX  
PT Novel polynucleotides useful for the treatment of various conditions  
PT including wounds and cancer -  
XX  
PS Claim 1; Page 204; 235pp; English.  
XX  
CC The invention relates to novel nucleic acid sequences derived from rat  
CC dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,  
CC and mouse embryonic skin, keratinocyte stem cells and transit amplifying  
CC cells. Polypeptides of the invention may be used to treat inflammation,  
CC cancer and neurological diseases. The proteins may be used to stimulate  
CC the growth and motility of keratinocytes, to inhibit the growth of  
CC cancer cells, to modulate angiogenesis and tumour vascularisation, to  
CC modulate skin inflammation, to modulate epithelial cell growth and to  
CC inhibit binding of HIV-1 to leukocytes. The invention may also be used  
CC to treat growth and developmental defects, skin wounds and hair follicle  
CC disorders. Sequences AAZ61606-261832 represent cDNA sequences derived  
CC from several mouse, rat or human skin cell types. Sequences  
CC AAZ61606-261649, AAZ61725-261765, AAZ61802-261811 and AAZ61826 encode  
CC proteins with an N-terminal signal sequence, indicating that the proteins  
CC are secreted. Sequences AAZ61650-261668, AAZ61766-261780, AAZ61812-261817  
CC and AAZ61827-261829 encode proteins with one or more putative  
CC transmembrane domains.  
XX  
SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;

Query Match 0.9%; Score 32; DB 21; Length 1837;  
Best Local Similarity 100.0%; Pred. No. 9.1e-05;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 1177  
|||||  
Db 942 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 973

## RESULT 5

AAC99736  
ID AAC99736 standard; cDNA; 1837 BP.

XX AC AAC99736;

XX DT 08-MAR-2001 (first entry)

XX DE Skin cell cDNA, SEQ ID NO: 350.

XX KW Mouse; skin cell; cytostatic; antiinflammatory; anti-HIV;

XX KW neotropic; neuroprotective; vulnerary; immunomodulatory; vaccine;

XX KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;

XX KW inflammation; neurological disease; ss.

XX OS Mus sp.

XX PN WO200069884-A2.

XX PD 23-NOV-2000.

XX PF 15-MAY-2000; 2000WO-NZ00075.

XX PR 14-MAY-1999; 99US-0312283.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PI Watson JD, Strachan L, Sleeman M, Kumble KD, Murison JG;

XX WPI; 2001-007495/01.

XX DR New isolated polynucleotide used in the identification of genetic  
PT disorders and encoding polypeptides used for treating inflammatory  
PT disease, cancer and neurological diseases -

XX PS Claim 1; Page 272-273; 352pp; English.

XX CC The present polynucleotide encodes a polypeptide which is expressed in  
CC mammalian skin cells. The polypeptide is useful for stimulating  
CC keratinocyte growth and motility, inhibiting the growth of cancer cells,  
CC modulating angiogenesis, inhibiting angiogenesis and vascularisation of  
CC tumours, modulating skin inflammation, stimulating the growth of  
CC epithelial cells, inhibiting the binding of human immunodeficiency virus  
CC (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and  
CC neurological diseases. The polynucleotide can be used as a marker, in  
CC the identification of genetic disorders, and for the design of  
CC oligonucleotides for examining expression patterns.

XX SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;

Query Match 0.9%; Score 32; DB 22; Length 1837;  
Best Local Similarity 100.0%; Pred. No. 9.1e-05;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 1177  
|||||  
Db 942 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 973

## RESULT 6

ABL34888

ID ABL34888 standard; cDNA; 1837 BP.

XX ABL34888;

XX AC ABL34888;

XX " "

XX DT

XX DE 04-APR-2002 (first entry)

XX DE Murine cDNA isolated from skin cells SEQ ID NO: 350.

XX DE

XX KW Human; rat; mouse; skin cell; skin wound; cancer; growth defect;

XX KW developmental defect; inflammatory disease; dermatological; vulnerary;

XX KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;

XX KW ss.

XX OS Mus sp.

XX PN WO200190357-A1.

XX PD 29-NOV-2001.

XX PF 24-MAY-2001; 2001WO-NZ00099.

XX PR 24-MAY-2000; 2000US-206650P.

XX PR 25-JUL-2000; 2000US-221232P.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PI Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;

XX WPI; 2002-122020/16.

XX DR New polynucleotides and polypeptides encoded by the polynucleotides  
XX isolated from skin cells, useful for treating skin wounds, cancers,  
XX growth and developmental defects, inflammatory diseases, or for  
XX modulating immune responses -

XX PS Claim 1; Page 227-228; 466pp; English.

XX CC The present invention provides the protein and coding sequences of cDNAs  
XX isolated from human, murine and rat skin cell libraries. The sequences  
XX can be used in the development of therapeutic agents useful in the  
XX treatment of skin diseases, including skin wounds, cancer, growth  
XX defects, developmental defects and inflammatory diseases. The proteins  
XX have important roles in the induction of hair growth, cell proliferation  
XX and cell-cell interaction, in maintaining tissue integrity, in wound  
XX healing and in modulating immune responses. The present sequence is a  
XX cDNA of the invention.

XX SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;

Query Match 0.9%; Score 32; DB 24; Length 1837;  
Best Local Similarity 100.0%; Pred. No. 9.1e-05;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 1177

|||||

Db 942 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 973

## RESULT 7

AAS92463

ID AAS92463 standard; cDNA; 1422 BP.

XX AC AAS92463;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #28267.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

```
XX 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG28276.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 28267; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) is useful for generating antibodies against it, detecting or
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS94197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 0.7%; Score 25; DB 23; Length 1422;
XX Best Local Similarity 100.0%; Pred. No. 0.21;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1236 CTGCCAGAAATGGAGGCACATGTGTT 1260
Db 258 CTGCCAGAAATGGAGGCACATGTGTT 282
RESULT 8
AAS92464
ID AAS92464 standard; cDNA; 11344 BP.
XX AAS92464;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #28268.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
```

```
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG28277.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 28268; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) is useful for generating antibodies against it, detecting or
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 0.7%; Score 25; DB 23; Length 11344;
XX Best Local Similarity 100.0%; Pred. No. 0.2;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1236 CTGCCAGAAATGGAGGCACATGTGTT 1260
Db 575 CTGCCAGAAATGGAGGCACATGTGTT 599
RESULT 9
AAC05374
ID AAC05374 standard; cDNA; 394 BP.
XX AAC05374;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 9449.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
```

```
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX PN WPI; 2000-500381/45.
XX DR
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 9449; 71pp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX SQ Sequence 394 BP; 102 A; 84 C; 96 G; 112 T; 0 other;

Query Match          0.7%; Score 23; DB 21; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 CAGCCCTGCCAGATGGAGGCAC 1253
Db 282 CAGCCCTGCCAGATGGAGGCAC 304

RESULT 10
AAD28942
ID AAD28942 standard; cDNA; 6728 BP.
AC AAD28942;
XX 07-MAY-2002 (first entry)
DE Human MOL1b cDNA.
KW Secreted molecule; MOL1b protein; MOLX; cardiomyopathy; atherosclerosis;
KW diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;
KW liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;
KW cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;
KW immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;
KW HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain;
KW haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; muscular disease; stress;
KW ocular disease; growth disorder; depression; epilepsy; contraceptive;
KW vulnary; osteopathic; haemostatic; tranquiliser; antidepressant;
KW analgesic; vasotropic; hypotensive; gene therapy; chromosome 1; ss.
XX OS Homo sapiens.
XX FH Key
FH FT 5'UTR 1..12 Location/Qualifiers
FH FT /*tag= a
FH FT 13..4434
FH FT /*tag= b
FH FT /*product= "Human MOL1b protein"
FH FT sig_peptide 13..87
FH FT /*tag= c
FH FT mat_peptide 88..4431
FH FT /*tag= d

/product= "Human MOL1b mature protein"
4435..6728
/*tag= e

3'UTR
WO200206339-A2.
24-JAN-2002.
03-JUL-2001; 2001WO-US21249.
03-JUL-2000; 2000US-215854P.
03-JUL-2000; 2000US-215856P.
03-JUL-2000; 2000US-215902P.
07-JUL-2000; 2000US-216585P.
07-JUL-2000; 2000US-216586P.
07-JUL-2000; 2000US-216722P.
17-JUL-2000; 2000US-218622P.
17-JUL-2000; 2000US-218992P.
27-JUL-2000; 2000US-221285P.
14-FEB-2001; 2001US-268734P.
(CURA-) CURAGEN CORP.
Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;
Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;
Alsbrook J, Lepley DM, Shen L, Burgess CE, Shinkets RA;
Padigaru M;
WPI; 2002-155038/20.
P-PSDB; AAE18208.
Nucleic acids encoding secreted polypeptides, designated MOLX
polypeptides, useful for treating a MOLX-associated disorder, e.g.
cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -
Claim 3; Page 14-15; 223pp; English.
The patent discloses nucleic acid sequences encoding novel secreted
molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL
protein where X is an integer from 1 to 8). Sequences of the invention
are useful for treating or preventing a MOLX-associated disorder in
humans. They are useful for treating or preventing cardiomyopathy,
atherosclerosis and disorders related to cell signal processing and
metabolic pathway modulation. The MOLX antibodies are useful for
treating or preventing diabetes and disorders related to cell signal
processing and metabolic pathway modulation. MOLX sequences are useful
for the treatment or diagnosis of other MOLX-associated disorders, e.g.
chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,
scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral
thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune
diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,
bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,
ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
ocular disease, muscular diseases, growth disorders, loss of libido,
stress, depression, pain and epilepsy. They are useful for preventing
chemotherapy side effects and as contraceptives. Sequences of the
invention are also useful for gene therapy. The present sequence
is a cDNA encoding human Notch-like protein, MOL1b. MOL1b gene is
localised on chromosome 1.
XX SQ Sequence 6728 BP; 1571 A; 1647 C; 1668 G; 1842 T; 0 other;

Query Match          0.7%; Score 23; DB 24; Length 6728;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 CAGCCCTGCCAGATGGAGGCAC 1253
Db 3592 CAGCCCTGCCAGATGGAGGCAC 3614

RESULT 11
```

AAD28941  
ID AAD28941 standard; cDNA; 7410 BP.

XX  
AC AAD28941;

XX  
DT 07-MAY-2002 (first entry)

XX  
DE Human MOLLA cDNA.

XX  
KW Secreted molecule; MOLLA protein; MOLX: cardiomyopathy; atherosclerosis;  
KW diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;  
KW liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;  
KW cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;  
KW immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;  
KW HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain;  
KW haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; muscular disease; stress;  
KW ocular disease; growth disorder; depression; epilepsy; contraceptive;  
KW vulnary; osteopathic; haemostatic; tranquiliser; antidepressant;  
KW analgesic; vasotropic; hypotensive; gene therapy; ss.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT CDS 1..7410

FT /tag= a

FT /product= "Human MOLLA protein"

FT sig\_peptide 1..75

FT /tag= b

FT mat\_peptide 76..7407

FT /tag= c

FT /product= "Human MOLLA mature protein"

XX  
FT WO200206339-A2.

XX  
PN 24-JAN-2002.

XX  
PD 03-JUL-2001; 2001WO-US21249.

XX  
PF 03-JUL-2000; 2000US-215854P.

XX  
PR 03-JUL-2000; 2000US-215856P.

XX  
PR 03-JUL-2000; 2000US-215902P.

XX  
PR 07-JUL-2000; 2000US-216585P.

XX  
PR 07-JUL-2000; 2000US-216586P.

XX  
PR 17-JUL-2000; 2000US-216722P.

XX  
PR 17-JUL-2000; 2000US-218622P.

XX  
PR 17-JUL-2000; 2000US-218992P.

XX  
PR 27-JUL-2000; 2000US-221285P.

XX  
PR 14-FEB-2001; 2001US-268734P.

XX  
PA (CURA-) CURAGEN CORP.

XX  
PI Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;

XX  
PI Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;

XX  
PI Alsobrook J, Lepley DM, Shen L, Burgess CE, Shinkets RA;

XX  
PI Fadigar M;

XX  
XX WPI; 2002-155038/20.

XX  
DR P-PSDB; AAE18207.

XX  
XX Nucleic acids encoding secreted polypeptides, designated MOLX

XX  
PT polypeptides, useful for treating a MOLX-associated disorder, e.g.

XX  
PT cardiomyopathy, atherosclerosis, diabetes and metabolic disorders

XX  
XX Claim 3; Page 9-11; 223pp; English.

XX  
XX The patent discloses nucleic acid sequences encoding novel secreted

XX  
CC molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL

XX  
CC protein where x is an integer from 1 to 8). Sequences of the invention

XX  
CC are useful for treating or preventing a MOLX-associated disorder in

XX  
CC humans. They are useful for treating or preventing cardiomyopathy,

XX  
CC atherosclerosis and disorders related to cell signal processing and

XX  
CC metabolic pathway modulation. The MOLX antibodies are useful for

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

XX  
 SQ Sequence 1851 BP; 379 A; 558 C; 574 G; 340 T; 0 other;

Query Match 0.7%; Score 22; DB 22; Length 1851;  
 Best Local Similarity 100.0%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;  
 Matches 22; Conservative 0;

QY 1190 CCACCTGCTACAGGACCACCTG 1211  
 |||||  
 DB 1727 CCACCTGCTACAGGACCACCTG 1748

## RESULT 13

AAK53119/c  
 ID AAK53119 standard; cDNA; 1851 BP.

XX  
 AC AAK53119;

XX  
 DT 06-NOV-2001 (first entry)

XX  
 DE Human polynucleotide SEQ ID NO 2648.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.

XX  
 OS Homo sapiens.

XX  
 PN WO200157190-A2.

XX  
 PD 09-AUG-2001.

XX  
 PF 05-FEB-2001; 2001WO-US04098.

XX  
 PR 03-FEB-2000; 2000US-0496914.

XX  
 PR 27-APR-2000; 2000US-0560875.

XX  
 PR 20-JUN-2000; 2000US-0598075.

XX  
 PR 19-JUL-2000; 2000US-0620325.

XX  
 PR 01-SEP-2000; 2000US-0654936.

XX  
 PR 15-SEP-2000; 2000US-0663561.

XX  
 PR 20-OCT-2000; 2000US-0693325.

XX  
 PR 30-NOV-2000; 2000US-0728422.

XX  
 PA (HYSE-) HYSEQ INC.

XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX  
 DR WPI; 2001-476283/51.

XX  
 DR P-PSDB; AAM79986.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -

XX  
 PS Claim 1; Page 4903; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

XX  
 SQ Sequence 1851 BP; 340 A; 574 C; 558 G; 379 T; 0 other;

Query Match 0.7%; Score 22; DB 22; Length 1851;  
 Best Local Similarity 100.0%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;  
 Matches 22; Conservative 0;

QY 1190 CCACCTGCTACAGGACCACCTG 1211  
 |||||  
 DB 125 CCACCTGCTACAGGACCACCTG 104

## RESULT 14

AAK59627

ID AAK59627 standard; cDNA; 498 BP.

XX  
 AC AAK59627;

XX  
 DT 02-FEB-2001 (first entry)

XX  
 DE Human secreted protein cDNA sequence #4.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; ss.

XX  
 OS Homo sapiens.

XX  
 PN WO2000056767-A1.

XX  
 PD 28-SEP-2000.

XX  
 PF 16-MAR-2000; 2000WO-US06828.

XX  
 PR 19-MAR-1999; 99US-0125358.

XX  
 PR 08-DEC-1999; 99US-0169616.

XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;

XX  
 DR WPI; 2000-602216/57.

XX  
 DR P-PSDB; AAB34442.

XX Nucleic acid molecules encoding human secreted proteins, used in  
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 PT Parkinson's diseases and cancers -

XX  
 PS Claim 1; Page 316; 384p; English.

XX The invention relates to the isolation of genes AAC59624-C59669 encoding  
 CC XX human secreted proteins AAB34439-B34484. The genes can be used to  
 CC generate fusion proteins by linking to the gene for the human  
 CC immunoglobulin G Fc portion (AAC59615) for increasing the stability of  
 CC the fusion protein as compared to the human protein only. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The

CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)  
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

XX SQ Sequence 498 BP; 109 A; 184 C; 93 G; 112 T; 0 other;

Query Match 0.6%; Score 20; DB 21; Length 498;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 CGGATCCTCAGGATCTGTTTC 1848

|||||

Db 360 CGGATCCTCAGGATCTGTTTC 379

#### RESULT 15

AAS64374/c

ID AAS64374 standard; cDNA; 972 BP.

XX AC AAS64374;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #178.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX PT WPI; 2001-639362/73.

XX PT P-PSDB; ABG00187.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity

XX PS Claim 1; SEQ ID No 178; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 972 BP; 252 A; 186 C; 263 G; 271 T; 0 other;

Query Match 0.6%; Score 20; DB 23; Length 972;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3049 AATGCCCCAGCAGAGGCCTTT 3068

|||||

Db 563 AATGCCCCAGCAGAGGCCTTT 544

Search completed: May 5, 2003, 05:07:01

Job time : 521 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:59:44 ; Search time 926 Seconds  
(without alignments)  
4306.664 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgctgcgcgctgac.....tgtccacctgaaggtcttc 3375

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 746064 seqs, 590810554 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3375	100.0	3375	9 US-09-930-020A-1	Sequence 1, Appl
2	32	0.9	1837	9 US-10-152-661-350	Sequence 350, App
3	32	0.9	1837	9 US-09-866-050A-350	Sequence 350, App
4	20	0.6	531	9 US-09-918-995-19791	Sequence 19791, A
5	20	0.6	1120	9 US-09-938-842A-2842	Sequence 2842, Ap
6	20	0.6	3638	12 US-10-044-090-493	Sequence 493, App
7	19	0.6	123	10 US-09-864-761-29368	Sequence 29368, A
8	19	0.6	297	10 US-09-983-965-5125	Sequence 5125, Ap
9	19	0.6	505	10 US-09-864-761-12802	Sequence 12802, A
10	19	0.6	952	10 US-09-822-849A-583	Sequence 583, App
11	19	0.6	3047	10 US-09-818-143-7	Sequence 7, Appl
12	19	0.6	32367	9 US-10-158-160A-14	Sequence 14, Appl
13	19	0.6	172637	10 US-09-805-458A-3	Sequence 3, Appl
14	18	0.5	384	10 US-09-920-300A-1005	Sequence 1005, Ap
15	18	0.5	384	12 US-10-033-528-1005	Sequence 1005, Ap
16	18	0.5	401	9 US-09-946-807-1125	Sequence 1125, Ap
17	18	0.5	401	10 US-09-795-668-1125	Sequence 1125, Ap
18	18	0.5	401	10 US-09-795-686-1125	Sequence 1125, Ap
19	18	0.5	405	9 US-10-060-036-386	Sequence 386, App

20	18	0.5	408	9 US-09-918-995-17637	Sequence 17637, A
21	18	0.5	411	9 US-09-918-995-31656	Sequence 31656, A
22	18	0.5	439	9 US-09-918-995-27118	Sequence 27118, A
23	18	0.5	448	9 US-09-918-995-20362	Sequence 20362, A
24	18	0.5	455	9 US-10-060-036-191	Sequence 191, App
25	18	0.5	463	9 US-09-918-995-27657	Sequence 27657, A
26	18	0.5	469	9 US-09-918-995-10605	Sequence 10605, A
27	18	0.5	484	9 US-09-918-995-2281	Sequence 2281, Ap
28	18	0.5	491	9 US-09-918-995-1244	Sequence 1244, Ap
29	18	0.5	560	9 US-09-764-891-2501	Sequence 2501, Ap
30	18	0.5	600	10 US-09-864-761-8401	Sequence 8401, Ap
31	18	0.5	645	10 US-09-995-598-93	Sequence 93, Appl
32	18	0.5	717	10 US-09-815-242-7765	Sequence 7765, Ap
33	18	0.5	962	10 US-09-452-239-23	Sequence 23, Appl
34	18	0.5	1023	10 US-09-452-239-25	Sequence 25, Appl
35	18	0.5	1031	10 US-09-840-795-3	Sequence 3, Appl
36	18	0.5	1504	9 US-10-245-103-79	Sequence 79, Appl
37	18	0.5	1504	9 US-10-245-107-79	Sequence 79, Appl
38	18	0.5	1504	9 US-10-245-143-79	Sequence 79, Appl
39	18	0.5	1504	9 US-10-245-771-79	Sequence 79, Appl
40	18	0.5	1504	9 US-10-245-851-79	Sequence 79, Appl
41	18	0.5	1504	9 US-10-245-883-79	Sequence 79, Appl
42	18	0.5	1504	9 US-10-237-535-79	Sequence 79, Appl
43	18	0.5	1504	9 US-10-238-183-79	Sequence 79, Appl
44	18	0.5	1504	9 US-10-238-283-79	Sequence 79, Appl
45	18	0.5	1504	9 US-10-238-370-79	Sequence 79, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-930-020A-1

; Sequence 1, Application US/09930020A

; Publication No. US20030077568A1

; GENERAL INFORMATION:

; APPLICANT: Gish, Kurt C.

; APPLICANT: Mack, David H.

; APPLICANT: Willson, Keith E.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Colorectal Cancer. Compositions

; TITLE OF INVENTION: and Methods of Screening for Colorectal Cancer

; FILE OF INVENTION: Modulators

; FILE REFERENCE: 018501-0031000US

; CURRENT APPLICATION NUMBER: US/09/930,020A

; CURRENT FILING DATE: 2001-08-14

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 3375

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: CBF9

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (328)..(2751)

; OTHER INFORMATION: CBF9

; US-09-930-020A-1

Query Match 100.0%; Score 3375; DB 9; Length 3375;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACAGTGTTCGCGCTCCACCGCTCGGAGCTGGGTGACCGCTAGAGTGAAGTACTT 60

|||||

Db 1 GACAGTGTTCGCGCTCCACCGCTCGGAGCTGGGTGACCGCTAGAGTGAAGTACTT 60

|||||

..oy 61 TTTTATTGACAGCTGGCGCGATGCGGCTTTAAAAACGCGGGGCTCTATGCACCTC 120

|||||

Db 61 TTTTATTGACAGCTGGCGCGATGCGGCTTTAAAAACGCGGGGCTCTATGCACCTC 120

|||||

121 CTTGGGGTAGTTCCTCCGACCTACGCGGGTCGGGTGCGTGGCCCTCTCCAGAGAG 180  
121 CTTGGGGTAGTTCCTCCGACCTACGCGGGTCGGGTGCGTGGCCCTCTCCAGAGAG 180  
181 ACAACAGGTGTCCACAGTGGACGCGCCCGCGGCGCCCTCTCTGATCCGCTAGCG 240  
181 ACAACAGGTGTCCACAGTGGACGCGCCCGCGGCGCCCTCTCTGATCCGCTAGCG 240  
241 CCCCTGGCCGAGCGCGCGCGGCTGTGTAGTAGAGCGCGCGGCGCCAGCGCTGG 300  
241 CCCCTGGCCGAGCGCGCGGCTGTGTAGTAGAGCGCGCGGCGCCAGCGCTGG 300  
301 TCCCGGCTCTCTTCGTTTATATCAACATGCGCCCTCTCTGTGTGAGGCGCTGTGT 360  
301 TCCCGGCTCTCTTCGTTTATATCAACATGCGCCCTCTCTGTGTGAGGCGCTGTGT 360  
361 GTTTTCTCTGTTTCCAGAGTGGCCCATCTCTCCCTCTCCAGGAAGTCCATGTAACAAA 420  
361 GTTTTCTCTGTTTCCAGAGTGGCCCATCTCTCCCTCTCCAGGAAGTCCATGTAACAAA 420  
421 GAAACCATCGGGAAGATTTTCAGTGCAGCAAAATGATGTGCTCGGCTGCAGTGGAC 480  
421 GAAACCATCGGGAAGATTTTCAGTGCAGCAAAATGATGTGCTCGGCTGCAGTGGAC 480  
481 ATCATGTTTCTGTAGTGGTCTACAGGTCGGGAAAGGAGCTTTGAAAGTCCAAAG 540  
481 ATCATGTTTCTGTAGTGGTCTACAGGTCGGGAAAGGAGCTTTGAAAGTCCAAAG 540  
541 CACTTTGCCATCACGTCTGTGACGTCGTGGACATCAGCCCGAGAGGTCAGAGTGGGA 600  
541 CACTTTGCCATCACGTCTGTGACGTCGTGGACATCAGCCCGAGAGGTCAGAGTGGGA 600  
601 GCATTCAGTTTCCACTCTCTCATCTGGAATCCCTTGGATTCATTTCAACCCAA 660  
601 GCATTCAGTTTCCACTCTCTCATCTGGAATCCCTTGGATTCATTTCAACCCAA 660  
661 CAGGAGTGAAGCAAGATCAGAGAGTGGTTTCAAGGAGGCGCGAGAGCGGAA 720  
661 CAGGAGTGAAGCAAGATCAGAGAGTGGTTTCAAGGAGGCGCGAGAGCGGAA 720  
721 CTTGCTCTGAAATPACCTTCTGCACAGAGGTTGCTTGGAGGCAAGAAATGCTTGTGCC 780  
721 CTTGCTCTGAAATPACCTTCTGCACAGAGGTTGCTTGGAGGCAAGAAATGCTTGTGCC 780  
781 CAGATCCTCATCATCTGCTACGTAGGGAAGTCCAGGGGATGTGCACTGCCATCCAAG 840  
781 CAGATCCTCATCATCTGCTACGTAGGGAAGTCCAGGGGATGTGCACTGCCATCCAAG 840  
841 CAGCTGAAGGAAAGGGGTGCTACTGTTTGTGCTGTTGGGGTCAGGTTTCCAGGTGGGAG 900  
841 CAGCTGAAGGAAAGGGGTGCTACTGTTTGTGCTGTTGGGGTCAGGTTTCCAGGTGGGAG 900  
901 GAGCTGATGCACTGGCCAGCGCTAGAGGCGAGTACGTGCTTGGCTGAGCAGGTG 960  
901 GAGCTGATGCACTGGCCAGCGCTAGAGGCGAGTACGTGCTTGGCTGAGCAGGTG 960  
961 GAGGATGCCAACACGGCTTTTCAGACCTCAGCAGTTCGCGCATCTCCAGCGCC 1020  
961 GAGGATGCCAACACGGCTTTTCAGACCTCAGCAGTTCGCGCATCTCCAGCGCC 1020  
1021 ACCCCAGAGTGCAGGTCAGAGCTACCCCTGTGAGCAGAGGAGCTGGAGATGTCGCG 1080  
1021 ACCCCAGAGTGCAGGTCAGAGCTACCCCTGTGAGCAGAGGAGCTGGAGATGTCGCG 1080  
1081 GAGTTCGCTGGCAATGCCCATGCTGGAGGATCGCGCGGACCCCTTGGGTGCTGGCT 1140  
1081 GAGTTCGCTGGCAATGCCCATGCTGGAGGATCGCGCGGACCCCTTGGGTGCTGGCT 1140  
1141 GCACATGTCCTTCTACAGCTGGAAGAGTGTCTTAAACCCACCTGCGACCTGCTAC 1200  
1141 GCACATGTCCTTCTACAGCTGGAAGAGTGTCTTAAACCCACCTGCGACCTGCTAC 1200

1201 AGACACACCTGCCAGGCGCCCTGTGACTGCGAGCCCTGCCAGAAATGGAGCAGATGTGT 1260  
1201 AGACACACCTGCCAGGCGCCCTGTGACTGCGAGCCCTGCCAGAAATGGAGCAGATGTGT 1260  
1261 CAGAAAGGAGTGCAGCGCTACAGTGCCTTGCCTGCGCTGGCCCTTGGAGGAGGCTAAC 1320  
1261 CAGAAAGGAGTGCAGCGCTACAGTGCCTTGCCTGCGCTGGCCCTTGGAGGAGGCTAAC 1320  
1321 TGTGCCCTGAAGTGCAGCTGGAATGCAGGCTGCACCTCTCTTCTGCTGGACAGCTCT 1380  
1321 TGTGCCCTGAAGTGCAGCTGGAATGCAGGCTGCACCTCTCTTCTGCTGGACAGCTCT 1380  
1381 GCGGACACACCTTGCAGCGCTTCTGCGGGCCAAAGTCTTCTGTAAGCGGTTGTGCGG 1440  
1381 GCGGACACACCTTGCAGCGCTTCTGCGGGCCAAAGTCTTCTGTAAGCGGTTGTGCGG 1440  
1441 GCGCTGCTGAGGAGACTCTCGGGCCGAGTGGGTGTCGACATACAGCAGGAGCTG 1500  
1441 GCGCTGCTGAGGAGACTCTCGGGCCGAGTGGGTGTCGACATACAGCAGGAGCTG 1500  
1501 CTGTTGGCTGCTGCTGGGGAGTACCAGGATGTGCTGACCTGCTTCTGAGGCTTCGAT 1560  
1501 CTGTTGGCTGCTGCTGGGGAGTACCAGGATGTGCTGACCTGCTTCTGAGGCTTCGAT 1560  
1561 GGCATTCCCTTCCGTTGGTGGCCCGGCTGACGGGAGTGCCTTGGCGGAGCGGCGAG 1620  
1561 GGCATTCCCTTCCGTTGGTGGCCCGGCTGACGGGAGTGCCTTGGCGGAGCGGCGAG 1620  
1621 CGTGGCTTCCGAGGCGCCAGGACAGGCGGCGGCTGACGGGAGTGCCTTGGCGGAGCGGCGAG 1680  
1621 CGTGGCTTCCGAGGCGCCAGGACAGGCGGCGGCTGACGGGAGTGCCTTGGCGGAGCGGCGAG 1680  
1681 CTCACGTAGTACACCTCCGAGGATGAGTTGGGGGCGGCTGACGGGAGTGCCTTGGCGGAGCGGCGAG 1740  
1681 CTCACGTAGTACACCTCCGAGGATGAGTTGGGGGCGGCTGACGGGAGTGCCTTGGCGGAGCGGCGAG 1740  
1741 GAGCTCTCTCTGCTGCTGAGGCTGAGGCTGCGGCGGCTGAGGAGTGCCTTGGCGGAGCGGCGAG 1800  
1741 GAGCTCTCTCTGCTGCTGAGGCTGAGGCTGCGGCGGCTGAGGAGTGCCTTGGCGGAGCGGCGAG 1800  
1801 GGCAGCCCAAGAGCATGTGATGCTTACTCGGATCTCAGGATCTGTTCAACCAATCCCT 1860  
1801 GGCAGCCCAAGAGCATGTGATGCTTACTCGGATCTCAGGATCTGTTCAACCAATCCCT 1860  
1861 GAGCTCAGGGGAGTGTGACCGCGGCGGCGGCTGAGGCTGCGGCGGAGTGCCTTGGCGGAGCGGCGAG 1920  
1861 GAGCTCAGGGGAGTGTGACCGCGGCGGCGGCTGAGGCTGCGGCGGAGTGCCTTGGCGGAGCGGCGAG 1920  
1921 CTCGCTCTCATGTTGACACCTCTGCTCAGTAGGGCCGAGAAATTTGCTCAGATGCGAG 1980  
1921 CTCGCTCTCATGTTGACACCTCTGCTCAGTAGGGCCGAGAAATTTGCTCAGATGCGAG 1980  
1981 AGCTTTGTGAGAGCTGTGCTCCCTTCCAGTTTGGTGAACCTCAGCTGACACAGGCTCGGC 2040  
1981 AGCTTTGTGAGAGCTGTGCTCCCTTCCAGTTTGGTGAACCTCAGCTGACACAGGCTCGGC 2040  
2041 CTGGTGGTGTATGGCAGGCTGACAGTGCCTTGGGCTGACACCAACCCAGGCTCGG 2100  
2041 CTGGTGGTGTATGGCAGGCTGACAGTGCCTTGGGCTGACACCAACCCAGGCTCGG 2100  
2101 GCTGCGATGCTGGGCGCATTAGCCAGGCGGCTTACTAGTGGGTGGGCTCAGCGCGC 2160  
2101 GCTGCGATGCTGGGCGCATTAGCCAGGCGGCTTACTAGTGGGTGGGCTCAGCGCGC 2160  
2161 ACCGCCCTGCTGCACATCTATGACAAAGTATGACCTCCAGAGGCTGCCGCGCTGGT 2220  
2161 ACCGCCCTGCTGCACATCTATGACAAAGTATGACCTCCAGAGGCTGCCGCGCTGGT 2220  
2221 GTCCCCAAAGCTGTGTGTGCTACAGGCGGAGGCGGCTGACAGGATGCAGCGCTTCCT 2280  
2221 GTCCCCAAAGCTGTGTGTGCTACAGGCGGAGGCGGCTGACAGGATGCAGCGCTTCCT 2280  
2281 GCCCAAGAGCTGAGGAACAATGTCATCTCTGTTGCTGGGCGCTGTCCTA 2340



Query Match 0.9%; Score 32; DB 9; Length 1837;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGCTCCTTCTACAGCTGGAAGAGAGTGTTC 1177  
Db 942 CTGCTCCTTCTACAGCTGGAAGAGAGTGTTC 973

## RESULT 4

US-09-918-995-19791/c  
; Sequence 19791, Application US/09018995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19791  
; LENGTH: 531  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(531)  
; OTHER INFORMATION: n - A, T, C or G

US-09-918-995-19791

Query Match 0.6%; Score 20; DB 9; Length 531;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 TGCCTGAGGCAGCAATGCT 771  
Db 114 TGCCTGAGGCAGCAATGCT 95

## RESULT 5

US-09-938-842A-2842  
; Sequence 2842, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRI1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 2842  
; LENGTH: 1120  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana

US-09-938-842A-2842

Query Match 0.6%; Score 20; DB 9; Length 1120;  
Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ATGTTGTTGAAAAGTTTGA 2884  
Db 394 ATGTTGTTGAAAAGTTTGA 413

## RESULT 6

US-10-044-090-493  
; Sequence 493, Application US/10044090  
; Patent No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090  
; CURRENT FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: PERL Program  
; SEQ ID NO 493  
; LENGTH: 3638  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. US20020137081A1 400203.1

US-10-044-090-493

Query Match 0.6%; Score 20; DB 12; Length 3638;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2534 GCAAGTGTGGGATGCTGG 2553  
Db 1203 GCAAGTGTGGGATGCTGG 1222

## RESULT 7

US-09-864-761-29368/c  
; Sequence 29368, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 29368  
;; LENGTH: 123  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AP001537.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98  
;; OTHER INFORMATION: NT HIT: AL163218.2, EVALUATE 2.00e-63  
;; OTHER INFORMATION: EST\_HUMAN HIT: AI807844.1, EVALUATE 1.00e-01  
;; OTHER INFORMATION: SWISSPROT HIT: P43403, EVALUATE 7.10e+00  
US-09-864-761-29368

Query Match 0.6%; Score 19; DB 10; Length 123;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2718 TACCTTCTGGAATCTGT 2736  
|||||  
Db 121 TACCTTCTGGAATCTGT 103

RESULT 8  
US-09-983-965-5125  
;; Sequence 5125, Application US/09983965  
;; Patent No. US20020137160A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Warren, Wesley C.  
;; APPLICANT: Tao, Nengbing  
;; APPLICANT: Byatt, John C.  
;; APPLICANT: Mathialagan, Nagappan  
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
;; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
;; FILE REFERENCE: 37-21(10297)C  
;; CURRENT APPLICATION NUMBER: US/09/983,965  
;; CURRENT FILING DATE: 2001-10-26  
;; PRIOR APPLICATION NUMBER: US 09/465,231  
;; PRIOR FILING DATE: 1999-12-15  
;; PRIOR APPLICATION NUMBER: US 60/113,678  
;; PRIOR FILING DATE: 1998-12-17  
;; NUMBER OF SEQ ID NOS: 5912  
;; SEQ ID NO 5125  
;; LENGTH: 297  
;; TYPE: DNA  
;; ORGANISM: Bos taurus  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: 33-LIB34-046-Q1-E1-A2  
US-09-983-965-5125

Query Match 0.6%; Score 19; DB 10; Length 297;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1488 CAGCAGGAGCTGCTGTG 1506  
|||||  
Db 38 CAGCAGGAGCTGCTGTG 56  
RESULT 9  
US-09-864-761-12802/c  
;; Sequence 12802, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; FILE REFERENCE: Aeomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 12802  
;; LENGTH: 505  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AP001537.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
US-09-864-761-12802

Query Match 0.6%; Score 19; DB 10; Length 505;  
Best Local Similarity 100.0%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2718 TACCTTCTGAAGTCTGT 2736  
|||||  
Db 224 TACCTTCTGAAGTCTGT 206

## RESULT 10

US-09-822-849A-583/c  
; Sequence 583, Application US/09822849A  
; Patent No. US20020045170A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakar  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6403  
; CURRENT APPLICATION NUMBER: US/09/822,849A  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/195,582  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 598  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 583  
; LENGTH: 952  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-849A-583

Query Match 0.6%; Score 19; DB 10; Length 952;

Best Local Similarity 100.0%; Pred. No. 15;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 CTGTGTTTCTCTGTTTCC 375  
|||||  
Db 173 CTGTGTTTCTCTGTTTCC 155

## RESULT 11

US-09-818-143-7/c  
; Sequence 7, Application US/09818143  
; Patent No. US20020019000A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Michael G.  
; APPLICANT: Volkmuth, Wayne  
; APPLICANT: Klingler, Tod M.  
; TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES  
; FILE REFERENCE: PB-0004 CIP  
; CURRENT APPLICATION NUMBER: US/09/818,143  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PERL Program  
; SEQ ID NO 7  
; LENGTH: 3047  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE: .  
; OTHER INFORMATION: 1656953CB1  
US-09-818-143-7

Query Match 0.6%; Score 19; DB 10; Length 3047;

Best Local Similarity 100.0%; Pred. No. 15;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 CTGTGTTTCTCTGTTTCC 375  
|||||  
Db 2257 CTGTGTTTCTCTGTTTCC 2239

## RESULT 12

US-10-158-160A-14  
; Sequence 14, Application US/10158160A  
; Publication No. US20030059805A1  
; GENERAL INFORMATION:  
; APPLICANT: RAPPOLD-HOERBRAND, GUDRUN  
; APPLICANT: RAO, ERCOLE  
; TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION  
; FILE REFERENCE: 108351-00004  
; CURRENT APPLICATION NUMBER: US/10/158,160A  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: 09/147,699  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: PCT/EP97/05355  
; PRIOR FILING DATE: 1997-09-29  
; PRIOR APPLICATION NUMBER: 60/027,633  
; PRIOR FILING DATE: 1995-10-01  
; PRIOR APPLICATION NUMBER: EP/97100583.0  
; PRIOR FILING DATE: 1997-01-16  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 32367  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-158-160A-14

Query Match 0.6%; Score 19; DB 9; Length 32367;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CCGAGCGCTGTCGCGCT 308  
|||||  
Db 12890 CCGAGCGCTGTCGCGCT 12908

## RESULT 13

US-09-805-458A-3/c  
; Sequence 3, Application US/09805458A  
; Patent No. US20020042100A1  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al  
; TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CLO00722  
; CURRENT APPLICATION NUMBER: US/09/805,458A  
; CURRENT FILING DATE: 2001-03-14  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 172637  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(172637)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-805-458A-3

Query Match 0.6%; Score 19; DB 10; Length 172637;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 CCTGGAGGCAGAAATGCTT 772  
|||||  
Db 122712 CCTGGAGGCAGAAATGCTT 122694

## RESULT 14

US-09-920-300A-1005/c

; Sequence 1005, Application US/09920300A  
; Patent No. US20020136728A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Gordon E.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.547  
; CURRENT APPLICATION NUMBER: US/09/920,300A  
; CURRENT FILING DATE: 2001-07-31  
; NUMBER OF SEQ ID NOS: 1789  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1005  
; LENGTH: 384  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-920-300A-1005

Query Match 0.5%; Score 18; DB 10; Length 384;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 959 TGGAGGATGCCACCAACG 976  
|  
Db 375 TGGAGGATGCCACCAACG 358

## RESULT 15

US-10-033-528-1005/c  
; Sequence 1005, Application US/10033528  
; Patent No. US20020131971A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Gordon E.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.547C1  
; CURRENT APPLICATION NUMBER: US/10/033,528  
; CURRENT FILING DATE: 2001-12-26  
; NUMBER OF SEQ ID NOS: 1896  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1005  
; LENGTH: 384  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-033-528-1005

Query Match 0.5%; Score 18; DB 12; Length 384;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 959 TGGAGGATGCCACCAACG 976  
|  
Db 375 TGGAGGATGCCACCAACG 358

Search completed: May 5, 2003, 07:54:36  
Job time : 1316 secs

**THIS PAGE BLANK (USPTO)**